Contribution à l'analyse statistique des données fontionnelles

Matthieu Saumard

Résumé

Dans cette thèse, nous nous intéressons aux données fonctionnelles.

La généralisation du modèle linéaire généralisé fonctionnel au modèle défini par des équations estimantes est étudiée. Nous obtenons un théorème du type théorème de la limite centrale pour l'estimateur considéré. Les instruments optimaux sont estimés, et nous obtenons une convergence uniforme des estimateurs.

Nous nous intéressons ensuite à différents tests en données fonctionnelles. Il s'agit de tests non-paramétriques pour étudier l'effet d'une covariable aléatoire fonctionnelle sur un terme d'erreur, qui peut être directement observé comme une réponse ou estimé à partir d'un modèle fonctionnel comme le modèle linéaire fonctionnel. Nous avons prouvé, pour pouvoir mettre en oeuvre les différents tests, un résultat de réduction de la dimension qui s'appuie sur des projections de la covariable fonctionnelle. Nous construisons des tests de non-effet et d'adéquation en utilisant soit un lissage par un noyau, soit un lissage par les plus proches voisins. Un test d'adéquation dans le modèle linéaire fonctionnel est proposé. Tous ces tests sont étudiés d'un point de vue théorique et pratique.

Abstract

In this thesis, we are interested in the functional data.

The problem of estimation in a model of estimating equations is studied. We derive a central limit type theorem for the considered estimator. The optimal instruments are estimated, and we obtain a uniform convergence of the estimators.

We are then interested in various testing with functional data. We study the problem of nonparametric testing for the effect of a random functional covariate on an error term which could be directly observed as a response or estimated from a functional model like for instance the functional linear model. We proved, in order to construct the tests, a result of dimension reduction which relies on projections of the functional covariate. We have constructed no-effect tests by using a kernel smoothing or a nearest neighbor smoothing. A goodness-of-fit test in the functional linear model is also proposed. All these tests are studied from a theoretical and practical perspective.

N° d'ordre : **13 ISAR 09 / D13-09**

Institut National des Sciences Appliquées de Rennes
20, Avenue des Buttes de Coësmes • CS 70839 • F-35708 Rennes Cedex 7
Tel : 02 23 23 82 00 - Fax : 02 23 23 83 96

Matthieu SAUMARD 2013

Thèse

THESE INSA Rennes
sous le sceau de l'Université européenne de Bretagne
pour obtenir le titre de
DOCTEUR DE L'INSA DE RENNES
Spécialité : Mathématiques et applications

présentée par
Matthieu Saumard
ECOLE DOCTORALE : *MATISSE*
LABORATOIRE : *IRMAR-INSA UMR CNRS 6625*

Contributions à l'analyse statistique des données fonctionnelles

Thèse soutenue le 23.05.2013
devant le jury composé de :

Pascal Sarda
Professeur à l'Université Paul Sabastier, Toulouse / président
Hervé Cardot
Professeur à l'Université de Bourgogne / rapporteur
André Mas
Professeur à l'Université Montpellier 2 / rapporteur
James Ledoux
Professeur à l'INSA de Rennes / examinateur
Valentin Patilea
Professeur à l'ENSAI / Directeur de thèse

Contributions à l’analyse statistique des données fonctionnelles

Matthieu Saumard

En partenariat avec

*Je dédie cette thèse à mon épouse, **Marianne**.*

Remerciements

Tout d'abord, je remercie vivement Valentin Patilea d'avoir dirigé ma thèse. Depuis le début de mon master II et tout au long de mon projet, il a fait preuve de bienveillance et d'attention à mon égard. Ses grandes qualités humaines ainsi que ses conseils m'ont permis de franchir les différents paliers jusqu'à la soutenance.

Je remercie chaleureusement les deux rapporteurs, Hervé Cardot et André Mas, d'avoir accepté de juger ce travail. Je remercie Pascal Sarda d'avoir bien voulu présider le jury, ainsi que James Ledoux d'avoir accepté de faire partie de la commission d'examen.

Je tiens également à remercier César Sanchez-Sellero, d'une part pour son essentielle collaboration à ce projet, d'autre part pour son accueil et sa sympathie durant mon séjour à Santiago de Compostella. Je remercie Wenceslao Manteiga de m'avoir parrainé pendant ce séjour. J'en profite pour remercier Maria et Vanda, qui ont contribué à rendre ce séjour agréable.

Je voudrais à présent remercier tous les membres du centre de mathématiques de l'INSA de Rennes, qui m'ont particulièrement bien accueilli durant ces années de thèse. J'ai grandement apprécié la richesse de nos échanges, ainsi que le soutien reçu dans toutes mes démarches. En outre, je suis heureux d'avoir pu tisser des liens privilégiés avec certains d'entre eux.

Je souhaite également remercier les doctorants de l'étage et ceux de la tour de maths, avec lesquels je partage et continuerai de partager d'excellents souvenirs.

Enfin, je remercie ma famille et mes amis de m'avoir accompagné, soutenu et parfois supporté tout au long de cette thèse.

Table des matières

Résumé. Dans cette thèse, nous nous intéressons aux données fonctionnelles. La généralisation du modèle linéaire généralisé fonctionnel au modèle défini par des équations estimantes est étudiée. Nous obtenons un théorème du type théorème de la limite centrale pour l'estimateur considéré. Les instruments optimaux sont estimés, et nous obtenons une convergence uniforme des estimateurs. Nous nous intéressons ensuite à différents tests en données fonctionnelles. Il s'agit de tests non-paramétriques pour étudier l'effet d'une covariable aléatoire fonctionnelle sur un terme d'erreur, qui peut être directement observé comme une réponse ou estimé à partir d'un modèle fonctionnel comme le modèle linéaire fonctionnel. Nous avons prouvé, pour pouvoir mettre en oeuvre les différents tests, un résultat de réduction de la dimension qui s'appuie sur des projections de la covariable fonctionnelle. Nous construisons des tests de non-effet et d'adéquation en utilisant soit un lissage par un noyau, soit un lissage par les plus proches voisins. Un test d'adéquation dans le modèle linéaire fonctionnel est proposé. Tous ces tests sont étudiés d'un point de vue théorique et pratique.

Mots Clés. Données fonctionnelles ; Équations estimantes ; Tests non-paramétriques ; Lissage par noyau ; Modèle de régression ; Modèle linéaire fonctionnel ; Test de non-effet ; Test d'adéquation.

Abstract. In this thesis, we are interested in the functional data. The problem of estimation in a model of estimating equations is studied. We derive a central limit type theorem for the considered estimator. The optimal instruments are estimated, and we obtain a uniform convergence of the estimators. We are then interested in various testing with functional data. We study the problem of nonparametric testing for the effect of a random functional covariate on an error term which could be directly observed as a response or estimated from a functional model like for instance the functional linear model. We proved, in order to construct the tests, a result of dimension reduction which relies on projections of the functional covariate. We have constructed no-effect tests by using a kernel smoothing or a nearest neighbor smoothing. A goodness-of-fit test in the functional linear model is also proposed. All these tests are studied from a theoretical and practical perspective.

Keywords. Functional data ; Estimating equations ; Nonparametric testing ; Kernel smoothing ; Regression model ; Functional linear model ; Test of no-effect, Goodness-of-fit test.

Chapitre 1

Introduction

1.1 Les données fonctionnelles

Les données provenant de diverses branches des sciences (économétrie, biologie, environnement, ...) sont de plus en plus collectées sous forme de courbes. Cela s'explique par le perfectionnement des outils de mesure et le progrès des outils informatiques, tant au niveau de la mémoire (problème du stockage) qu'au niveau de la rapidité d'exécution des programmes informatiques. Ainsi, de très grands ensembles de données peuvent être observés. La statistique des données fonctionnelles qui traite ce genre de données est donc un sujet d'étude de plus en plus important. Mathématiquement, une définition très générale des données fonctionnelles, voir les monographies de Ramsay et Silvermann (2005) et de Ferraty et Vieu (2006) pour de nombreux exemples, est celle-ci : une donnée fonctionnelle est une observation d'une variable aléatoire qui prend ses valeurs dans un espace de dimension infinie. Dans ce manuscrit, on considère uniquement des données fonctionnelles à valeurs dans l'espace de fonctions réelles de carré intégrable sur l'intervalle $[0, 1]$ que nous notons $L^2[0, 1]$, bien que les nouvelles méthodologies décrites puissent s'adapter à des espaces plus généraux.

1.2 Des modèles fonctionnels

1.2.1 Le modèle linéaire fonctionnel

Je commence par la présentation du modèle linéaire fonctionnel avec la situation la plus courante, c'est-à-dire quand les réponses sont scalaires, pour plus de détails consulter Cardot et Sarda (2010). L'ensemble des données consiste en une séquence

(X_i, Y_i), $i = 1, \cdots, n$ où X_i est une fonction définie sur un intervalle I de $\mathbb{R}$ et $Y_i \in \mathbb{R}$. L'hypothèse minimale sur les courbes X_i est qu'elles sont de carré intégrable sur I. On prend généralement $[0, 1]$ pour I afin de simplifier les développements. On considère donc $L^2([0, 1])$ que l'on munit de son produit scalaire naturel $\langle f, g \rangle = \int_0^1 f(t)g(t)dt$ et de la norme induite. Cet ensemble est alors un espace de Hilbert. Les courbes aléatoires sont supposées avoir un second moment fini :

$$\mathbb{E}(||X_i||^2) < +\infty$$

On peut alors définir l'opérateur de covariance Γ :

$$\forall u \in L^2([0, 1]),\ \Gamma u = \mathbb{E}\left(\langle X_i - \mathbb{E}(X_i), u \rangle (X_i - \mathbb{E}(X_i))\right) \tag{1.2.1}$$

On voit aisément que Γ est un opérateur intégral de noyau la fonction covariance de X_i :

$$\forall u \in L^2([0, 1]), \forall t \in [0, 1],\ \Gamma u(t) = \int_0^1 \mathrm{Cov}(X_i(s), X_i(t)) u(s) ds$$

L'opérateur Γ est positif, symétrique, Hilbert-Schmidt et donc compact. On rappelle que les valeurs propres de Γ, $\lambda_j, j = 1, 2, \ldots$ sont strictement positives et que 0 est le seul point d'accumulation du spectre. On peut maintenant appliquer une décomposition de Karhunen-Loève sur la courbe X_i, en notant $v_j, j = 1, 2, \ldots,$ la séquence de fonctions propres associées aux valeurs propres λ_j, rangées dans l'ordre décroissant :

$$\forall t \in [0, 1],\ X_i(t) - \mathbb{E}(X_i(t)) = \sum_{j=1}^{+\infty} \xi_j v_j(t)$$

où $\xi_j, j = 1, 2, \ldots,$ sont des variables aléatoires centrées non-corrélées dont la variance est égale à λ_j :$\mathbb{E}(\xi_j) = 0$ et $\mathbb{E}(\xi_j \xi_{j'}) = \lambda_j \delta_{jj'}$

Enfin, nous pouvons écrire le modèle :

$$Y_i = \alpha + \int_0^1 \beta(t) X_i(t) dt + \epsilon_i,\ i = 1, \ldots, n,$$

où α est une constante et $\beta \in L^2([0, 1])$ est une fonction paramètre, ϵ_i est un échantillon i.i.d. de variables aléatoires non-corrélées avec X_i : $\mathbb{E}(\epsilon_i) = 0$ et $\mathbb{E}(X_i(t)\epsilon_i) = 0$ pour t p.s. dans $[0, 1]$. L'intérêt principal réside en l'estimation du paramètre fonctionnel β, dont découle celle de la constante. On a

$$\begin{aligned} \mathbb{E}(X_i - \mathbb{E}(X_i)(Y_i - \mathbb{E}(Y_i)) &= \mathbb{E}(X_i - \mathbb{E}(X_i) \int_0^1 \beta(t)(X_i(t) - \mathbb{E}(X_i)(t))dt \\ &= \Gamma \beta \end{aligned}$$

En identifiant $\mathbb{E}(X_i - \mathbb{E}(X_i)(Y_i - \mathbb{E}(Y_i))$ et l'opérateur de covariance croisée Δ, on a donc

$$\Delta = \Gamma\beta \tag{1.2.2}$$

L'équation (1.2.2) est une version continue des équations normales que satisfait le vecteur des coefficients de régression de la régression linéaire multivariée usuelle. L'existence de β n'est alors pas aussi simple que dans le cas multivarié. En effet, l'opérateur Γ n'admet pas d'inverse continu. Par conséquent, une solution β ne peut pas se déduire directement de (1.2.2). D'abord, le paramètre fonctionnel peut être identifié seulement dans l'espace orthogonal au noyau de Γ : si β satisfait (1.2.2) et si β_1 est dans le noyau de Γ ($\Gamma\beta_1 = 0$) alors $\beta + \beta_1$ satisfait aussi (1.2.2).

Considérons $v_j, j = 1, 2, \ldots$, un système orthonormal complet de fonctions propres, on peut écrire $\beta = \sum_{j=1}^{+\infty}\langle\beta, v_j\rangle v_j$. Par (1.2.2), on obtient

$$\langle\mathbb{E}((X_i - \mathbb{E}(X_i))(Y_i - \mathbb{E}(Y_i))), v_j\rangle = \lambda_j\langle\beta, v_j\rangle, \quad j = 1, 2, \ldots$$

Maintenant, on cherche une solution dans la fermeture de $Im(\Gamma)$ où on fait l'hypothèse, sans perte de généralité, que le noyau de Γ est réduit à zéro. On obtient alors la décomposition pour β :

$$\beta = \sum_{j=1}^{+\infty} \frac{\langle\mathbb{E}((X_i - \mathbb{E}(X_i))(Y_i - \mathbb{E}(Y_i))), v_j\rangle}{\lambda_j} v_j = \sum_{j=1}^{+\infty} \frac{\mathbb{E}(\xi_j(Y_i - \mathbb{E}(Y_i)))}{\lambda_j} v_j,$$

et la fonction β est dans $L^2[0,1]$ si et seulement si la condition suivante est satisfaite :

$$\sum_{j=1}^{+\infty} \frac{(\mathbb{E}(\xi_j(Y_i - \mathbb{E}(Y_i))))^2}{\lambda_j^2} < +\infty$$

Cette condition assure l'existence et l'unicité d'une solution β.

L'opérateur de covariance Γ étant inconnu, l'estimation de β est faite au moyen d'une méthode de régularisation combinée avec une estimation de Γ. Il existe alors deux classes de procédures d'estimation. La première s'appuie sur une méthode de projection sur un espace de dimension finie dont la dimension croît avec la taille de l'échantillon. La méthode la plus populaire ici est la régression fonctionnelle en composantes principales. L'autre classe de procédures est basée sur une décomposition du paramètre fonctionnel dans une base de fonctions de $L^2[0,1]$, combinée avec la minimisation d'une version empirique de (1.2.2) avec l'addition d'un terme de pénalisation qui a pour effet de régulariser la solution. La base la plus populaire est celle des fonctions splines.

Je décris maintenant la régression en composantes principales fonctionnelle. La classe des estimateurs β basés sur une ACP fonctionnelle consiste en une régression des moindres carrés ordinaires (OLS) des réponses Y_i sur les vecteurs des composantes principales, $(\langle\hat{\phi}_1, X_i\rangle, \ldots, \langle\hat{\phi}_K, X_i\rangle)'$, où $\hat{\phi}_1, \ldots, \hat{\phi}_K$ sont les fonctions propres associées aux K plus grandes valeurs propres de l'opérateur de covariance empirique Γ_n basée sur l'échantillon $X_1, \ldots, X_n$. Dans cette définition K est un entier positif définissant la dimension de l'espace projeté qui joue le même rôle qu'un paramètre fenêtre.

L'estimateur FPCR $\hat{\beta}$ de β est défini comme :

$$\hat{\beta} = \sum_{j=1}^{K} \frac{\Delta_n \hat{\phi}_j}{\hat{\lambda}_j} \hat{\phi}_j$$

où Γ_n est l'opérateur de covariance empirique,

$$\Gamma_n \cdot = \frac{1}{n} \sum_{i=1}^{n} \langle X_i, \cdot\rangle X_i$$

et Δ_n est l'opérateur de covariance croisée,

$$\Delta_n = \frac{1}{n} \sum_{i=1}^{n} X_i Y_i \quad .$$

Les contributions dans ce domaine sont Cardot et al (1999, 2003), Cai et Hall (2006), Hall et Hosseini-Nasab (2006, 2009), Cardot et al (2007), Hall et Horowitz (2007) et Reiss et Ogden (2007).

1.2.2 Le modèle linéaire fonctionnel généralisé

Le modèle linéaire fonctionnel généralisé ou modèle de quasi-vraisemblance fonctionnel a été étudié par Müller et Stadtmüller (2005). Nous sommes toujours dans le cadre d'un échantillon i.i.d. (Y_i, X_i), $i = 1, \ldots, n$, où la variable aléatoire X appartient à $L^2[0,1]$ et Y est une variable aléatoire réelle. Nous supposons connaître une fonction de lien ρ. De plus nous avons une fonction variance $\sigma^2(\cdot)$ qui est strictement positive. Le modèle linéaire fonctionnel généralisé est déterminé par une fonction paramètre $\theta(\cdot)$ qui est supposée appartenir à $L^2[0,1]$. Ce modèle s'écrit :

$$Y_i = \rho\left(\alpha + \int_0^1 \theta(t) X_i(t) dt\right) + e_i, \quad i = 1, \ldots, n, \qquad (1.2.3)$$

où

$$\mathbb{E}\left(e|X\right) = 0,$$

$$\mathrm{Var}(e|X) = \sigma^2(\mu)$$

où $\mu = \rho\left(\alpha + \int_0^1 \theta(t)X_i(t)dt\right)$. Soit ρ_j, $j = 1, 2, \ldots$, une base orthonormale de $L^2[0,1]$, X et θ se décomposent suivant cette base

$$X = \sum_{j=1}^{\infty} \epsilon_j \rho_j, \quad \theta = \sum_{j=1}^{\infty} \theta_j \rho_j.$$

avec ϵ_j des variables aléatoires et les coefficients θ_j. Il en résulte immédiatement que

$$\int_0^1 \theta(t)X(t)dt = \sum_{j=1}^{\infty} \theta_j \epsilon_j.$$

Müller et Stadtmüller approchent le modèle initial (1.2.3) par une série de modèles dont le nombre de prédicteurs est tronqué à $p = p_n$ et la dimension p_n croît vers l'infini quand le nombre de données n tend vers l'infini. Ainsi la suite de modèles considérés est

$$Y_i = \rho\left(\alpha + \sum_{j=1}^{p} \theta_j \epsilon_j^{(i)}\right) + e'_i. \tag{1.2.4}$$

En notant $\theta_0 = \alpha$, on peut estimer le paramètre vecteur inconnu $\theta^T = (\theta_0, \ldots, \theta_p)$ en résolvant l'équation “score”

$$U(\theta) = 0,$$

où en notant $\epsilon^{(i)T} = (\epsilon_0^{(i)}, \ldots, \epsilon_p^{(i)})$, $\nu_i = \sum_{j=0}^{p} \theta_j \epsilon_j^{(i)}$, $\mu_i = \rho(\nu_i)$, la fonction score U, à valeurs vectorielles, est définie par

$$U(\theta) = \sum_{i=1}^{n} (Y_i - \mu_i)\rho'(\nu_i)\epsilon^{(i)}\sigma^2(\mu_i).$$

On peut résoudre cette équation en la réécrivant à l'aide de matrice voir l'article de Müller et Stadtmüller (2005). En notant la solution $\hat{\theta}^T = (\hat{\theta}_0, \ldots, \hat{\theta}_p)$, on obtient la fonction estimée

$$\hat{\theta}(t) = \hat{\theta}_0 + \sum_{j=1}^{p} \hat{\theta}_j \rho_j(t).$$

On peut aussi citer l'approche de Cardot et Sarda (2005). Les auteurs estiment le coefficient fonctionnel du modèle via une vraisemblance pénalisée dans un contexte d'approximation par des splines. L'identifiabilité du modèle est discuté. Le taux de convergence L^2 de l'estimateur est donné sous des hypothèses de régularité du paramètre fonctionnel. Un récent travail de Dou, Pollard et Zhou (2012) établit les taux minimax optimaux de convergence pour l'estimation du paramètre fonctionnel. Ces taux dépendent de l'opérateur de covariance (la décroissance des valeurs propres) et de la régularité du paramètre fonctionnel.

1.3 Les tests d'hypothèses

Une contribution majeure de cette thèse se situe dans une nouvelle approche de construction de tests d'hypothèses en données fonctionnelles. En reprenant une idée de réduction de la dimension due à Lavergne et Patilea (2008) et en l'adaptant aux données fonctionnelles, l'hypothèse nulle des tests de non-effet et d'adéquation est réécrite suivant une projection des covariables sur une classe suffisamment riche d'éléments non-aléatoires d'hypersphère. On peut alors baser les tests sur une statistique qui découle soit d'un lissage de cette projection par un noyau **univarié** comme décrit dans le chapitre 3, et qui donc ne souffre pas de l'effet de la grande dimension, soit d'une approche par plus proches voisins présentée dans le chapitre 5.

1.3.1 Tests en données fonctionnelles

Je présente dans cette partie trois tests en données fonctionnelles. Le premier test du à Cardot et al. (2003) traite du modèle linéaire fonctionnel. Il s'agit de tester une hypothèse de non-effet qui est la nullité de l'opérateur de régression (ou la nullité de la fonction paramètre du modèle). Le second test que je présente ici est le test d'Horváth et Reeder (2011) dans le modèle fonctionnel de régression quadratique. Les auteurs proposent de tester l'importance du terme non-linéaire dans le modèle. Dans leur procédure de test, ils utilisent l'analyse en composantes principales fonctionnelles. Le troisième test auquel je fais référence est le test de Delsol et Al. (2010) qui est un test non-paramétrique. Nous pouvons aussi citer l'article de Cuesta-Albertos et al. (2006) qui propose un test d'adéquation à des modèles paramétriques. Ils utilisent une projection aléatoire de la variable X qui appartient à un espace de Hilbert $\mathcal{H}$ sur un élément aléatoire de $\mathcal{H}$. Ils génèrent k éléments aléatoires de $\mathcal{H}$ et leur statistique de test est le maximum de k Kolmogorov-Smirnov statistiques.

Test dans le modèle linéaire fonctionnel

Soient X une fonction aléatoire de carré intégrable sur $[0, 1]$ et Y une variable aléatoire réelle, le modèle s'écrit

$$Y = \int_0^1 \phi(t)X(t)dt + \epsilon.$$

Cardot et Al. (2003) étudient le test d'hypothèse suivant :

$$H_0 : \quad \phi = 0$$

contre l'alternative

$$H_1 : \quad \phi \neq 0.$$

Ils réécrivent l'hypothèse nulle en utilisant l'opérateur de covariance croisée Δ, qui est défini par

$$\Delta x = \int_0^1 \mathbb{E}[X(t)Y]x(t)dt.$$

En effet, tester "$\phi = 0$" est équivalent à tester "$\Delta = 0$". D'où, la nouvelle écriture de H_0,

$$H_0 : \quad \Delta = 0.$$

Notons $(\lambda_j), j = 1, \ldots$ la suite décroissante des valeurs propres de l'opérateur Γ et $(V_j), j = 1, \ldots$ une suite orthonormale de fonctions propres associées à ces valeurs propres, ainsi que leurs versions empiriques $(\hat{\lambda}_j), j = 1, \ldots$ et $(\hat{V}_j), j = 1, \ldots$. Définissons maintenant

$$\hat{A}_n(\cdot) = \sum_{j=1}^{p_n} \hat{\lambda}_j^{-1/2} \langle \hat{V}_j, \cdot \rangle \hat{V}_j,$$

alors la statistique de test proposée est

$$D_n = \frac{1}{\hat{\sigma}_n^2} \|\sqrt{n} \Delta_n \hat{A}_n\|^2,$$

où $\hat{\sigma}_n^2$ est un estimateur de la variance σ^2 de ϵ. D_n suit approximativement, quand n est grand, une loi du χ^2 à p_n degrées de liberté. On peut aussi considérer le test basé sur la statistique suivante :

$$T_n = \frac{1}{\sqrt{p_n}} \left(\frac{1}{\hat{\sigma}_n^2} \|\sqrt{n} \Delta_n \hat{A}_n\|^2 - p_n \right).$$

Test dans le modèle fonctionnel de régression quadratique

La régression fonctionnelle quadratique a été étudiée par Yao et Müller (2010). Le modèle est le suivant, en supposant la variable X centrée,

$$Y_n = \mu + \int_0^1 k(t) X_n dt + \int_0^1 \int_0^1 h(s,t) X_n(s) X_n(t) dt ds + \epsilon_n.$$

Pour tester l'importance du terme quadratique, l'hypothèse nulle est :

$$H_0 : \quad h(s,t) = 0,$$

contre l'alternative

$$H_1 : \quad h(s,t) \neq 0.$$

L'idée du test de Horváth et Reeder (2011) est de décomposer $h(\cdot,\cdot)$ dans la base orthonormale formée par les fonctions propres de l'opérateur de covariance du processus prédicteur et de fixer un entier positif p qui sert de troncature.
En notant les fonctions propres et les valeurs propres $(v_i(t), \lambda_i)$, $i = 1, \ldots,$ de l'opérateur de covariance de X, on peut décomposer h et k dans cette base :

$$h(s,t) = \sum_{i=1}^{\infty} a_{i,i} v_i(s) v_i(t) + \sum_{i=1}^{\infty} \sum_{j=i+1}^{\infty} a_{i,j} (v_j(s) v_i(t) + v_i(s) v_j(t)),$$

et

$$k(t) = \sum_{i=1}^{\infty} b_i v_i(t).$$

Le modèle peut se réécrire en projetant sur l'espace engendré par $\{v_1, \ldots, v_p\}$:

$$Y_n = \mu + \sum_{i=1}^{p} b_i \langle X_n, v_i \rangle + \sum_{i=1}^{p} \sum_{j=i}^{p} (2 - 1_{\{i=j\}}) a_{i,j} \langle X_n, v_i \rangle \langle X_n, v_j \rangle + \epsilon_n^*,$$

mais ne connaissant pas μ et $v_i(t)$, on les estime par leurs versions empiriques que l'on note $\bar{X}(t) = \frac{1}{N} \sum_{n=1}^{N} X_n(t)$ et $\hat{v}_i(t)$. Les fonctions propres correspondant à une seule valeur propre sont uniquement déterminées à un signe près : $\hat{c}_i$. On peut écrire le modèle sous forme matricielle :

$$Y = \hat{Z} \begin{pmatrix} \tilde{A} \\ \tilde{B} \\ \mu \end{pmatrix} + \epsilon^{**},$$

où $Y = (Y_1, Y_2, \ldots, Y_N)^T$, $\tilde{A} = \text{vech}(\{\hat{c}_i \hat{c}_j a_{i,j} (2 - 1_{\{i=j\}}), 1 \leq i \leq j \leq p\}^T)$, $\tilde{B} = (\hat{c}_1 b_1, \hat{c}_2 b_2, \ldots, \hat{c}_p b_p)^T$, $\epsilon^{**} = (\epsilon_1^{**}, \ldots, \epsilon_N^{**})^T$ et

$$\hat{Z} = \begin{pmatrix} \hat{D}_1^T & \hat{F}_1^T & 1 \\ \hat{D}_2^T & \hat{F}_2^T & 1 \\ \cdot & \cdot & \cdot \\ \hat{D}_N^T & \hat{F}_N^T & 1 \end{pmatrix}$$

avec

$$\hat{D}_n = \text{vech}\left(\{ \langle \hat{v}_i, X_n \rangle \langle \hat{v}_j, X_n \rangle, 1 \leq i \leq j \leq p \}^T \right),$$
$$\hat{F}_n = \left(\langle X_n, \hat{v}_1 \rangle, \langle X_n, \hat{v}_2 \rangle, \ldots, \langle X_n, \hat{v}_p \rangle \right)^T.$$

Les auteurs estiment $\tilde{A}$, $\tilde{B}$ et μ en utilisant un estimateur des moindres carrés :

$$\begin{pmatrix} \hat{A} \\ \hat{B} \\ \hat{\mu} \end{pmatrix} = \left(\hat{Z}^T \hat{Z} \right)^{-1} \hat{Z}^T Y.$$

Soient

$$\hat{G} = \frac{1}{N}\sum_{n=1}^{N}\hat{D}_n\hat{D}_n^T,$$

$$\hat{M} = \frac{1}{N}\sum_{n=1}^{N}\hat{D}_n,$$

et

$$\hat{\tau}^2 = \frac{1}{N}\sum_{n=1}^{N}\hat{\epsilon}_n^2,$$

où $\hat{\epsilon}_n$ sont les résidus sous H_0 à savoir :

$$\hat{\epsilon}_n = Y_n - \hat{\mu} - \sum_{i=1}^{p}\hat{b}_i\langle X_n, \hat{v}_i\rangle - \sum_{i=1}^{p}\sum_{j=i}^{p}(2 - 1_{\{i=j\}})\hat{a}_{i,j}\langle X_n, \hat{v}_i\rangle\langle X_n, \hat{v}_j\rangle.$$

La statistique de test est la suivante :

$$U_N = \frac{N}{\hat{\tau}^2}\hat{A}^T\left(\hat{G} - \hat{M}\hat{M}^T\right)\hat{A}.$$

Sous certaines conditions et sous l'hypothèse H_0, U_N converge en loi vers un χ^2 à r degrés de liberté où $r = p(p+1)/2$ qui est la dimension du vecteur $\hat{A}$. Les auteurs obtiennent que sous H_1, leur statistique de test diverge.

Test de structure sur données fonctionnelles

Considérons le modèle de régression suivant :

$$Y_i = r(X_i) + \epsilon_i$$

où les variables (X_i, Y_i) pour i de 1 à n sont indépendantes. Pour simplifier, je traite le cas de l'hypothèse nulle de la forme :

$$H_0 : \quad \mathbb{P}\left(r(X) = r_0(X)\right) = 1$$

contre la suite d'alternatives

$$H_1 : \quad \|r - r_0\|_{L^2(wdP_X)} \geq \tau_n,$$

où w est une fonction poids et P_X est la loi de X qui est connue. Le cas où r appartient à une famille plus générale qu'un singleton est aussi traité dans l'article de Delsol et al. (2010). Le test est construit à partir de la statistique suivante :

$$T_n = \int\left(\sum_{i=1}^{n}(Y_i - r_0(X_i))K\left(\frac{d(x, X_i)}{h_n}\right)\right)^2 w(x)dP_X(x),$$

où K est un noyau univarié, $d(\cdot,\cdot)$ est une semi-métrique et h_n est un paramètre fenêtre. En normalisant correctement cette statistique, les auteurs obtiennent la normalité asymptotique de leur statistique de test ainsi que la consistance du test. Définissons

$$T_{1,n} = \int \sum_{i=1}^{n} K^2 \left(\frac{d(X_i, x)}{h_n} \right) \epsilon_i^2 w(x) dP_X(x)$$

et

$$T_{2,n} = \int \sum_{1 \leq i \neq j \leq n} K \left(\frac{d(X_i, x)}{h_n} \right) K \left(\frac{d(X_j, x)}{h_n} \right) \epsilon_i \epsilon_j w(x) dP_X(x).$$

La statistique de test est

$$I_n = \frac{1}{\sqrt{\mathrm{Var}(T_{2,n})}} \left(T_n - \mathbb{E}[T_{1,n}] \right).$$

Cette statistique tend vers une loi normale centrée réduite sous H_0 et diverge vers $+\infty$ sous H_1.

1.3.2 Un test en données finies-dimensionnelles

Il existe une littérature abondante sur les tests non-paramétriques en données finies-dimensionnelles, notamment pour tester des modèles paramétriques : par exemple Härdle et Mammen (1993), Stute (1997), Escanciano (2006) et Guerre et Lavergne (2002). Le test que je mentionne maintenant est dû à Lavergne et Patilea (2008). J'ai choisi de le présenter car nous avons généralisé l'idée de réduction de dimension présentée dans ce test aux données fonctionnelles et suivi la procédure de test dans le chapitre 3.
L'hypothèse nulle est de la forme :

$$H_0 : \quad \mathbb{E}[U(\theta_0)|X] = 0,$$

où θ_0 est un paramètre inconnu qui doit être estimé et $X \in \mathbb{R}^q$. Dans le cas d'une régression paramétrique, $U(\theta)$ est de la forme $U(\theta) = Y - \mu(X;\theta)$, où $\mu(\cdot,\cdot)$ appartient à une famille paramétrique et θ appartient à un sous-ensemble de $\mathbb{R}^d$. Le test repose sur une idée de réduction de la dimension qui montre que tester l'espérance conditionnelle sachant X est équivalent à tester l'espérance conditionnelle sachant $X'\beta$, quelque soit β de norme 1. Soit

$$Q(\theta, \beta) = \mathbb{E}\left[U(\theta) \mathbb{E}[U(\theta)|X'\beta] f_\beta(X'\beta) \right],$$

où $f_\beta(\cdot)$ est la densité de la variable aléatoire $X'\beta$. Définissons

$$Q_n(\theta, \beta) = \frac{1}{n(n-1)} \sum_{j \neq i} U_i(\theta) U_j(\theta) \frac{1}{h} K_h((X_i - X_j)'\beta),$$

où $U_i(\theta) = Y_i - \mu(X_i;\theta)$ et $K_h(\cdot) = K(\cdot/h)$, où $K(\cdot)$ est un noyau et h une fenêtre. Ils doivent choisir β et pour cela ils utilisent une méthode par pénalisation :

$$\hat{\beta}_n = \arg\max_{\|\beta\|=1}\{nh^{1/2}Q_n(\hat{\theta}_n,\beta) - \alpha_n 1_{\{\beta\neq\beta_0\}}\},$$

où $\hat{\theta}_n$ est un estimateur de θ_0, β_0 est choisi par l'utilisateur et α_n, $n \geq 1$, est une suite de nombres réels positifs décroissants vers 0. La statistique de test est :

$$T_n = nh^{1/2}\frac{Q_n(\hat{\theta}_n,\hat{\beta}_n)}{\hat{v}_n(\beta_0)},$$

où $\hat{v}_n^2(\beta)$ est un estimateur de la variance de $nh^{1/2}Q_n(\hat{\theta}_n,\beta)$. Les auteurs ont montré que cette statistique est asymptotiquement normale sous H_0, ainsi que la consistance du test.

1.4 Contributions des différents chapitres

On présente, ci-dessous, les différentes contributions de la thèse répertoriées par chapitre. Le chapitre 3 est l'objet d'un article soumis qui s'intitule "Projection-based nonparametric goodness-of-fit testing with functional covariates". L'article suivant issu du chapitre 5 s'intitule "Nonparametric testing for no-effect with functional responses and functional covariates". Ces deux articles ont été écrits en collaboration avec Valentin Patilea et César Sanchez-Sellero. Les deux autres chapitres font l'objet d'articles en cours.

Chapitre 2 : Nous proposons un nouveau modèle de données fonctionnelles, il s'agit d'un modèle dit de restrictions de moments conditionnels ou modèle dit d'équations estimantes pour des données fonctionnelles. Ce modèle est largement étudié dans le domaine de l'économétrie pour des données finies-dimensionnelles. Notre modèle généralise le modèle linéaire fonctionnel généralisé. Un estimateur du paramètre fonctionnel d'intérêt est proposé, et la convergence de l'estimateur vers la vraie fonction est démontrée. Nous proposons aussi une estimation des instruments optimaux.

Chapitre 3 : Une nouvelles méthode de réduction de la dimension pour des données fonctionnelles est présentée dans le lemme fondamental. Ce lemme nous permet de formuler de nouvelles statistiques de test pour tester l'effet d'une covariable fonctionnelle ou l'adéquation à des modèles paramétriques. Notre statistique de test est une forme quadratique basée sur un lissage par un noyau univarié et les valeurs critiques asymptotiques sont données par la loi normale centrée réduite. Le test est capable de détecter des alternatives non-paramétriques. Le terme d'erreur peut présenter de l'hétéroscédasticité de forme inconnue. Il est à souligner que la loi de covariable n'est pas supposée connue. Nous avons aussi contourné le problème des petites boules de probabilité, qui reste un sujet à développer. Les résultats de simulation confirment la théorie et montrent que les différents tests se comportent bien. Une application sur données réelles est aussi proposée.

Chapitre 4 : Nous étendons les résultats du chapitre 3 au cas d'une réponse fonctionnelle. Le test de non-effet pour une réponse fonctionnelle est étudié en détail. Comme précédemment, aucune hypothèse restrictive sur la loi des variables fonctionnelles n'est requise : les variables ne sont pas supposées gaussiennes, leurs lois ne sont pas connues. La preuve est similaire à celle du chapitre 3. Toutefois, la preuve de la convergence vers une gaussienne centrée réduite est modifiée. En effet, on utilise un théorème dû à Hall (1984).

Chapitre 5 : Une nouvelle approche pour le test de non-effet avec une réponse fonctionnelle est présentée. La statistique de test étudiée est une forme quadratique utilisant un lissage par les plus proches voisins et les valeurs critiques asymptotiques sont données par la loi normale centrée réduite. En utilisant un lissage par les plus proches voisins, on change complètement les preuves du chapitre 4. La preuve s'articule autour d'un théorème central limit de de Jong (1987) qui est une généralisation de Hall (1984) ; et l'utilisation permanente des propriétés des plus proches voisins. Cette nouvelle procédure ainsi que celle du chapitre 4 peuvent aussi s'appliquer au cas de réponse réelle. Le test est capable de détecter des alternatives non-paramétriques. Comme pour les autres tests, la loi de la covariable n'a pas besoin d'être connue.

Chapitre 2

Estimation par la méthode des moments généralisée pour des équations estimantes

2.1 Introduction

Plusieurs modèles économétriques impliquent des restrictions de moments conditionnels. Hansen (1982) a proposé pour estimer le paramètre d'intérêt dans ce type de modèle la méthode des moments généralisée (méthode G.M.M.). Ces modèles ont été étudiés par la suite par Newey (1993). Plusieurs autres méthodes, comme par exemple la vraisemblance empirique, ont été proposées depuis ; citons Kitamura, Tripathi, et Ahn (2004), Smith (2007) et Dominguez et Lobato (2004). Toutes ces méthodes s'appliquent à des variables finies-dimensionnelles.

Le modèle étudié le plus proche de celui que l'on considère est celui de Müller et Stadtmüller (2005). Ils étudient les modèles linéaires fonctionnels généralisés. L'idée principale est de réduire la dimension des variables fonctionnelles en les décomposant dans une base orthonormale et d'utiliser les p premières composantes pour se ramener à un modèle fini-dimensionnel, puis de faire tendre $p = p_n$ vers l'infini. Le modèle que l'on considère ici est une généralisation de ce modèle. Nous allons utiliser la même idée de réduction de la dimension. On utilise la méthode des moments généralisée pour estimer les p premières valeurs du paramètre d'intérêt fonctionnel dans la décomposition dans la base orthonormale. Cette fonction paramètre d'intérêt remplace le vecteur paramètre d'intérêt dans les modèles finis-dimensionnels de restrictions de moments conditionnels.

Nous obtenons un résultat asymptotique de consistance de l'estimateur vers la fonction paramètre d'intérêt, quand la dimension de troncature croît avec la dimension de l'échantillon. Le deuxième résultat est un résultat asymptotique limite (type théorème centrale limite) de la déviation entre le paramètre fonctionnel estimé et le vrai paramètre fonctionnel.

Le chapitre est organisé comme suit : La description du modèle, des notations et des considérations préliminaires se trouvent dans la Section 2. La procédure d'estimation du paramètre fonctionnel est décrite dans la Section 3. Les résultats asymptotiques sont regroupés dans la Section 4. Dans la Section 5, nous estimons les instruments optimaux dans un cas particulier. Une étude de simulation est présentée dans la Section 6. Enfin, les preuves des résultats sont détaillées à la fin du chapitre, Section 7.

2.2 Le modèle

On dispose de n observations i.i.d. $z_i = (x_i, y_i)$, $i = 1, \ldots, n$ issues du couple (x, y) où $x \in L^2[0,1]$ est une courbe aléatoire et y est une variable aléatoire réelle. On notera g une fonction connue à valeurs vectorielles de dimension $s \times 1$. Notre but est d'estimer un paramètre d'intérêt $\theta_0 \in L^2[0,1]$ définit par

$$\mathbb{E}\left[g(z,\theta) \mid x\right] = 0 \Longleftrightarrow \theta = \theta_0 \tag{2.2.1}$$

Nous allons nous cantonner au cas :

$$g((x,y),\theta) = \bar{g}\left(y, \int_0^1 x(t)\theta(t)dt\right)$$

Le modèle linéaire généralisé fonctionnel étudié par Müller et Stadtmüller (2005) peut s'écrire de la manière suivante :

$$y_i = \rho\left(\alpha + \int_0^1 \theta(t)x_i(t)dt\right) + e_i, \quad i = 1 \ldots n, \tag{2.2.2}$$

où ρ est une fonction à valeurs réelles et le terme d'erreur satisfait les conditions $\mathbb{E}[e_i \mid x] = 0$ et $\mathrm{Var}(e_i \mid x) = \sigma^2(\mu)$, et $\mu = \rho\left(\alpha + \int_0^1 \theta(t)x(t)dt\right)$. Le modèle (2.2.1) est une généralisation du modèle (2.2.2). En posant $\bar{g}(y, \int_0^1 x(t)\theta(t)dt) = y - \rho\left(\alpha + \int_0^1 \theta(t)x(t)dt\right)$, on retrouve les restrictions de moments conditionnels : $\mathbb{E}\left[\bar{g}(y, \int_0^1 x(t)\theta(t)dt) \mid x\right] = 0$. Le modèle (2.2.1) que nous étudions ici a donc aussi comme cas particulier le modèle linéaire fonctionnel. En effet, il est déjà un cas particulier du modèle linéaire généralisé fonctionnel (2.2.2) ; il suffit de prendre $\rho(z) = z$, $\forall z \in \mathbb{R}$ et ainsi nous obtenons

$$y_i = \alpha + \int_0^1 \theta(t)x_i(t)dt + e_i, \quad i = 1 \dots n,$$

qui est le modèle linéaire fonctionnel.

Un troisième exemple de modèle inclus dans le modèle d'équations estimantes (2.2.1) est le modèle de régression quantile. De plus, ce dernier modèle ne s'écrit pas sous la forme du modèle linéaire généralisé fonctionnel. Le modèle de régression quantile s'écrit :

$$\mathbb{P}\left(y \leq \int_0^1 x(t)\theta(t)dt|x\right) = \tau.$$

En posant $g(y,z) = 1_{\{y\leq z\}} - \tau$, nous retrouvons les restrictions de moments conditionnels, $\mathbb{E}\left[g(y, \int_0^1 x(t)\theta(t)dt) \mid x\right] = 0$.

Soit ρ_j, $j = 1, 2, \dots$, une base orthonormale de l'espace $L^2[0,1]$, on a donc $\int_0^1 \rho_j(t)\rho_k(t)dt = \delta_{jk}$. Ainsi $x(t)$ et $\theta(t)$ peuvent se décomposer en :

$$x(t) = \sum_{j=1}^{\infty} \epsilon_j \rho_j(t) \ , \ \theta(t) = \sum_{j=1}^{\infty} \theta_j \rho_j(t),$$

l'égalité étant prise au sens des fonctions $L^2[0,1]$, c'est-à-dire presque sûrement. Les coefficients ϵ_j et θ_j sont donnés par $\epsilon_j = \int x(t)\rho_j(t)dt$ et par $\theta_j = \int \theta(t)\rho_j(t)dt$. De l'orthonormalité de la base ρ_j, il vient immédiatement que

$$\int_0^1 x(t)\theta(t)dt = \sum_{j=1}^{\infty} \epsilon_j \theta_j.$$

On tronque alors cette somme aux $p = p_n$ premiers termes pour pouvoir se ramener à un modèle fini-dimensionnel. On note $x^{(p)}$ le vecteur de taille p et ϵ_j sa j-ème composante (notation générique pour désigner la troncation d'un élément de $L^2[0,1]$ en un vecteur de dimension p) :

$$x^{(p)} = \left(\epsilon_1 = \int_0^1 x(t)\rho_1(t)dt, \dots, \epsilon_p = \int_0^1 x(t)\rho_p(t)dt\right)'.$$

Nous rencontrons alors un problème d'approximation de modèle. En effet, nous ne sommes pas sûr de trouver $\mathbb{E}\left[g(y, \sum_{j=1}^{p} \epsilon_j \theta_{0,j}) \mid x^{(p)}\right] = 0$. Cependant, quand

$p = p_n \to +\infty$, nous pouvons écrire, sous de bonnes conditions :

$$\mathbb{E}\left[g(y, \sum_{j=1}^{p} \epsilon_j \theta_{0,j}) \mid x^{(p)}\right] = O_{\mathbb{P}}(r_n), \tag{2.2.3}$$

où r_n est un nombre réel qui tend vers zéro, qu'on définit ci-dessous. En effet, posons $U_p^0 = \sum_{j=1}^{p} \epsilon_j \theta_{0,j}$ et $V_p^0 = \sum_{j=p+1}^{\infty} \epsilon_j \theta_{0,j}$ de tel sorte qu'on ait $\int_0^1 x(t)\theta_0(t)dt = U_p^0 + V_p^0$, on obtient alors

$$\begin{aligned}
\mathbb{E}\left[g(y, U_p^0) \mid x^{(p)}\right] &= \mathbb{E}\left[g(y, U_p^0) - \mathbb{E}[g(y, U_p^0 + V_p^0) \mid x] \mid x^{(p)}\right] \\
&= \mathbb{E}\left[g(y, U_p^0) - g(y, U_p^0 + V_p^0) \mid x^{(p)}\right] \\
&= \int g(s, U_p^0) - g(s, U_p^0 + V_p^0) dF_{y|x^{(p)}}(s).
\end{aligned}$$

Sous les conditions $|g(s, u_1) - g(s, u_2)| \leq \Phi(s)|u_1 - u_2|$ et $\int \Phi(s) dF_{y|x^{(p)}}(s) \leq C$, $\forall p$, pour une constante C, nous pouvons majorer l'intégrande :

$$\begin{aligned}
\mathbb{E}\left(\mathbb{E}^2\left[g(y, U_p^0) \mid x^{(p)}\right]\right) &\leq \mathbb{E}\left[\left(\int \Phi(s) dF_{y|x^{(p)}}(s)\right)^2 (V_p^0)^2\right] \\
&\leq C^2 \mathbb{E}[(V_p^0)^2] \\
&\leq C^2 \sum_{j=p+1}^{\infty} \theta_{0,j}^2 \sum_{j=p+1}^{\infty} \sigma_j^2
\end{aligned}$$

où $\sigma_j^2 = \mathbb{E}[\epsilon_j^2]$. On peut donc prendre $r_n = \sqrt{\sum_{j=p+1}^{\infty} \theta_{0,j}^2 \sum_{j=p+1}^{\infty} \sigma_j^2}$. L'erreur d'approximation est directement reliée à $Var(V_p^0)$ et est contrôlée par la suite des variances de ϵ_j, et la valeur du vrai paramètre. La base ρ_j joue ici un rôle primordial, plus les premiers coefficients dans la décomposition sont informatifs, plus l'erreur d'approximation sera faible.

2.3 Estimation

Avant de pouvoir mettre en oeuvre la méthode, il faut ramener les restrictions de moments conditionnels à des restrictions de moments inconditionnels. On utilise pour cela une variable instrumentale $A(x)$ de dimension $r \times s$. De

$$\mathbb{E}\left[g\left(y, \int x\theta_0\right) \mid x\right] = 0,$$

on peut écrire
$$\mathbb{E}\left[A(x)g\left(y,\int x\theta_0\right)\right]=0.$$
Pour plus de simplicité, on utilise les instruments optimaux $B(x^{(p)})$ et dans un premier temps, on considère ces instruments optimaux connus. Dans cette partie, nous avons juste besoin de connaître la définition des instruments optimaux. Les propriétés ainsi que quelques exemples sont proposés dans la section 5. Rappelons la définition des instruments optimaux :

$$\begin{aligned}
B(x^{(p)}) &= D(x^{(p)})'\Omega^{-1}(x^{(p)}) \qquad \text{où} \\
D(x^{(p)}) &= \mathbb{E}[\frac{\partial g}{\partial\theta}(y,x^{(p)},\theta_0^{(p)})|x^{(p)}] \qquad \text{et} \\
\Omega(x^{(p)}) &= \mathbb{E}[g(y,x^{(p)},\theta_0^{(p)})g(y,x^{(p)},\theta_0^{(p)})'|x^{(p)}].
\end{aligned}$$

$D(x^{(p)})$ est une matrice de taille $s\times p$, $\Omega(x^{(p)})$ est une matrice de taille $s\times s$, et la matrice des instruments optimaux $B(x^{(p)})$ est donc de taille $p\times s$. Dans notre notation de la dérivée, nous faisons un abus de notation : en fait
$$\frac{\partial g}{\partial\theta}(y,x^{(p)},\theta_0^{(p)})=x'^{(p)}g_2'(y,\langle x^{(p)},\theta_0\rangle),$$
dans le cas où $s=1$ et g_2' est la dérivée de g par rapport à la deuxième variable. Nous faisons aussi un abus de notation dans l'expression $g(y,x^{(p)},\theta_0^{(p)})$ qui est égal à $g(y,\langle x^{(p)},\theta_0^{(p)}\rangle)$. Les instruments optimaux sont évidement inconnues, puisqu'ils dépendent de θ_0, mais nous les utilisons dans l'analyse statistique comme des quantités connues. Nous verrons dans la section 5 comment estimer ces quantités et les remplacer dans l'estimateur. Soit $\hat{P}$ une matrice aléatoire $p\times p$ semi-définie positive qui peut-être dans un premier temps égal à l'identité. Considérons l'estimateur
$$\hat{\theta}_n=\arg\min_{\theta\in\Theta}\hat{g}_n(\theta)'\hat{P}\hat{g}_n(\theta)$$
où
$$\hat{g}_n(\theta)=\frac{1}{n}\sum_{i=1}^n B(x_i^{(p)})g(y_i,\sum_{j=1}^p\theta_j\epsilon_j^i).$$
Et ϵ_j^i représente la j-ème coordonnée de x_i dans la base orthonormée. $\hat{\theta}_n$ est un vecteur de dimension p, que nous comparons à
$$\theta_0^{(p)}=\left(\theta_{0,1}=\int_0^1\theta_0(t)\rho_1(t)dt,\ldots,\theta_{0,p}=\int_0^1\theta_0(t)\rho_p(t)dt\right)'.$$
C'est l'objet de la section suivante.

2.4 Résultats asymptotiques

Introduisons tout d'abord quelques notations : on note $\|\cdot\|$ les normes quadratiques des vecteurs de $\mathbb{R}^p$ (bien qu'elles varient avec n), on note de plus $\xi_{k_1,k_1'} = \Lambda_{k_1,k_1'}$, où Λ est la matrice $(\mathbb{E}[D(x)'\Omega(x)^{-1}D(x)])^{-1}$ de taille $p \times p$; $\xi^{(1/2)}_{k_1,k_1'} = \Lambda^{-1/2}_{k_1,k_1'}$. Pour alléger les notations, on écrit $\tilde{g}^k = B(x_k^{(p)})g(y_k, \sum_{j=1}^p \theta_j \epsilon_j^k)$, qui est un vecteur de taille p, de plus $\tilde{g}_k$ est la k-ème coordonnée du vecteur $\tilde{g}^1$.

Nous sommes maintenant en mesure d'établir un résultat de consistance :

Lemma 4.1 *Sous les conditions*

(a) Θ *est compact.*

(b) soit Φ *une fonction telle que pour tout* $s \in \mathbb{R}$, $g(s, \cdot)$ *est Lipschitzienne :*

$$|g(s, u_1) - g(s, u_2)| \leq \Phi(s)|u_1 - u_2|.$$

(c) la fonction Φ *définie en* (b) *vérifie pour une constante* $C > 0$,

$$\int \Phi(s) dF_{y|x}(s) \leq C.$$

(d) condition d'uniformité sur les troncatures :

$$\sup_p \mathbb{E} \sup_\theta \|B(x^{(p)})g(y, \langle x^{(p)}, \theta^{(p)} \rangle)\| < \infty.$$

on a

$$||\hat{\theta}_n - \theta_0^{(p)}|| = o_{\mathbb{P}}(1).$$

Ce résultat porte sur la consistance de l'estimateur sur les coordonnées de $\theta_0(\cdot)$ dans la base ρ_j. Il ne s'agit pas d'un résultat sur l'estimateur fonctionnel vers la vraie fonction $\theta_0(\cdot)$. Nous nous appuyons sur la démonstration classique de la consistance de l'estimateur G.M.M. pour démontrer ce résultat. Les hypothèses données ici sont cependant plus fortes que dans le cas classique. Par exemple, pour l'hypothèse (d), on a besoin d'une uniformité en p, ce qui nous permet d'appliquer une loi des grands nombres uniformes.

Donnons quelques hypothèses supplémentaires pour établir le résultat central :

Assumption D

(a) La fonction connue g *est deux fois dérivable.*

(b) *La fonction g vérifie pour un $a \in]0,1]$*

$$\left\| \frac{\partial g}{\partial \theta}(z_i, b) - \frac{\partial g}{\partial \theta}(z_i, d) \right\|_2 \leq c(y_i) \|b - d\|^a \quad \forall i = 1, \ldots, n, \, \forall b, d,$$

et

$$\|B_j(x_i)\| |c(y_i)| \leq \Gamma(z_i), \, \forall j = 1, \ldots, p, \, avec \quad \mathbb{E}[\Gamma(z_i)] < \infty \, \forall i = 1, \ldots, n.$$

(c) *Conditions sur les coefficients : il existe α, β, γ tel que*

$$\theta_{0,j} \sim j^{-\alpha}, \quad \sigma_j \sim j^{-\beta}, \quad p = n^{\gamma},$$

avec $\alpha \geq 2$, $\beta \geq 2$, $0 < \gamma < 1/4$ et

$$1 - 2\gamma(\alpha + \beta - 1) + \frac{3}{2}\gamma < 0.$$

(d) *on a*

$$\sum_{k_1, k_2, k'_1, k'_2}^{p} \xi_{k_1, k'_1} \xi_{k_2, k'_2} \mathbb{E}[(\frac{\partial \tilde{g}^k}{\partial \theta})_{k_1, k_2} (\frac{\partial \tilde{g}^k}{\partial \theta})_{k'_1, k'_2}] = o(n/p^2) \quad , \forall k.$$

(e) *on a*

$$\sum_{k_1, k_2, k_3, k_4 = 1}^{p} \mathbb{E}[\tilde{g}_{k_1} \tilde{g}_{k_2} \tilde{g}_{k_3} \tilde{g}_{k_4}] \xi_{k_1, k_2} \xi_{k_3, k_4} = o(n/p^2).$$

(f) *on a*

$$\sum_{k_1, \cdots, k_8 = 1}^{p} \mathbb{E}[\tilde{g}_{k_1} \tilde{g}_{k_3} \tilde{g}_{k_5} \tilde{g}_{k_7}] \mathbb{E}[\tilde{g}_{k_2} \tilde{g}_{k_4} \tilde{g}_{k_6} \tilde{g}_{k_8}] \xi_{k_1, k_2} \xi_{k_3, k_4} \xi_{k_5, k_6} \xi_{k_7, k_8} = o(n^2 p^2).$$

Commentons ces hypothèses. Les conditions sur les coefficients nous permettent d'écrire $p = o(n^{1/4})$, $r_n \sim n^{-\gamma(\alpha+\beta-1)}$ et ainsi on obtient la condition $nr_n^2 p^{3/2} = o(1)$. Cette dernière condition est utilisée dans les preuves et est valide pour $\alpha = \beta = 2$ et $\gamma = 1/4$. Les hypothèses D$-(e) - (f)$ dans le cas du modèle linéaire fonctionnel généralisé décrit par Müller et Stadtmüller (2005) sont exactement les mêmes. C'est-à-dire que si on pose $\bar{g}(y, \int_0^1 x(t)\theta(t)dt) = y - \rho\left(\alpha + \int_0^1 \theta(t)x(t)dt\right)$, alors on retrouve les conditions M$-(3)$ et M$-(4)$ du théorème 4.1 de Müller et Stadtmüller (2005). L'hypothèse D$-(d)$ s'ajoute du fait que l'on considère un modèle d'équations estimantes. Nous pouvons maintenant formuler le résultat principal :

Theorem 4.2 *Sous les hypothèses D et sous les conditions du lemme 4.1, et en posant* Λ *tel que* $\Lambda = (\mathbb{E}[D(x)'\Omega(x)^{-1}D(x)])^{-1}$, *on obtient :*

$$\frac{n\left(\hat{\theta}_n - \theta_0^{(p)}\right)' \Lambda^{-1}\left(\hat{\theta}_n - \theta_0^{(p)}\right) - p}{\sqrt{2p}} \Rightarrow N(0,1)$$

Notons

$$\hat{\theta}_n(t) = \sum_{i=1}^{p} \hat{\theta}_{n,i}\rho_i(t),$$

l'estimateur fonctionnel de θ_0. Pour pouvoir comparer ces deux fonctions, il faut une distance dans $L^2[0,1]$. La distance usuelle ne convient pas, puisque la matrice Λ^{-1} intervient dans le théorème (4.2) sous la forme $\left(\hat{\theta}_n - \theta_0^{(p)}\right)' \Lambda^{-1}\left(\hat{\theta}_n - \theta_0^{(p)}\right)$ qui n'est pas la distance usuelle de $\mathbb{R}^p$. Soit, pour une fonction g réelle c'est-à-dire $s = 1$,

$$G(u,t) = \mathbb{E}\left[\frac{\mathbb{E}^2[g_2'(y, \int x\theta_0)|x]}{\mathbb{E}[g^2(y, \int x\theta_0)|x]} x(u)x(t)\right], \tag{2.4.4}$$

un noyau intégral, où la dérivée de g est prise sur la seconde variable. Sous la condition,

$$\mathbb{E}\left[\frac{\mathbb{E}^2[g_2'(y, \int x\theta_0)|x]}{\mathbb{E}[g^2(y, \int x\theta_0)|x]} \|x\|^2\right] < +\infty,$$

G est défini presque partout sur $[0,1]^2$. En effet, il suffit pour cela de montrer que G est dans $L^1\left([0,1]^2\right)$. Nous avons

$$\begin{aligned}\iint_{[0,1]^2} |G(u,t)|dudt &\leq \iint_{[0,1]^2} \mathbb{E}\left[\frac{\mathbb{E}^2[g_2'(y, \int x\theta_0)|x]}{\mathbb{E}[g^2(y, \int x\theta_0)|x]} |x(u)||x(t)|\right] dudt \\ &\leq \mathbb{E}\left[\frac{\mathbb{E}^2[g_2'(y, \int x\theta_0)|x]}{\mathbb{E}[g^2(y, \int x\theta_0)|x]} \|x\|^2\right] < +\infty\end{aligned}$$

Faisons correspondre à ce noyau intégral la distance dans $L^2[0,1]$ définie comme suit :

$$d_G^2(f,g) = \iint (f(u) - g(u))(f(t) - g(t))G(u,t)dtdu, \ \forall f, g \in L^2[0,1].$$

Nous pouvons aussi définir l'opérateur linéaire intégral A_G correspondant à G :

$$(A_G f)(t) = \int f(u)G(u,t)du.$$

Cet opérateur est diagonalisable dans $L^2[0,1]$ si $\int |G(u,t)|^2 dudt < \infty$. Notons $\{\rho_j^G, \lambda_j^G, j = 1, 2, \cdots\}$, les fonctions propres et les valeurs propres de l'opérateur A_G. La distance d_G peut s'exprimer comme

$$d_G^2(f,g) = \sum_k \lambda_k^G \left(f_{\rho^G,k} - g_{\rho^G,k}\right)^2,$$

où $f_{\rho^G,k}$ (respectivement $g_{\rho^G,k}$) est le k-ème coefficient dans la décomposition de f (respectivement g) dans la base ρ^G. Nous obtenons alors

$$d_G^2(\hat{\theta}_n(\cdot), \theta_0(\cdot)) = \left(\hat{\theta}_n^G - \theta_0^{(p)G}\right)' \Lambda^{-1} \left(\hat{\theta}_n^G - \theta_0^{(p)G}\right) + \sum_{j=p+1}^{\infty} \lambda_j^G \theta_{0,j}^2,$$

où tous les calculs sont effectués dans la base ρ^G. En effet,

$$\begin{aligned}
\delta_{j,k}\lambda_j^G &= \iint G(u,t)\rho_j^G(s)\rho_k^G(t)dudt \\
&= \iint \mathbb{E}\left[x(u)x(t)\frac{\mathbb{E}^2[g_2'(y, \int x\theta_0)|x]}{\mathbb{E}[g^2(y, \int x\theta_0)|x]}\right]\rho_j^G(u)\rho_k^G(t)dudt \\
&= \mathbb{E}\left[\epsilon_j\epsilon_k\frac{\mathbb{E}^2[g_2'(y, \int x\theta_0)|x]}{\mathbb{E}[g^2(y, \int x\theta_0)|x]}\right] \\
&= \mathbb{E}\left[\epsilon_j^G\epsilon_k^G\frac{\mathbb{E}^2[g_2'(y, \int x^{(p)}\theta_0)|x^{(p)}]}{\mathbb{E}[g^2(y, \int x^{(p)}\theta_0)|x^{(p)}]}\right] \\
&= \Lambda_{j,k}^{-1},
\end{aligned}$$

en remplaçant les ϵ_j par ϵ_j^G qui sont donnés par

$$\epsilon_j^G = \frac{\mathbb{E}[g_2'(y, \int x\theta_0)|x]}{\mathbb{E}[g_2'(y, \int x^{(p)}\theta_0)|x^{(p)}]}\frac{\sqrt{\mathbb{E}[g^2(y, \int x^{(p)}\theta_0)|x^{(p)}]}}{\sqrt{\mathbb{E}[g^2(y, \int x\theta_0)|x]}}\int x(t)\rho_j^G(t)dt,$$

Toutes ces considérations donnent le résultat suivant :

Corollary 4.3 *Si la fonction $\theta_0(\cdot)$ vérifie la propriété*

$$\sum_{j=p+1} \lambda_j^G \left(\int_0^1 \theta_0(t)\rho_j^G dt\right)^2 = o(\frac{\sqrt{p}}{n}), \tag{2.4.5}$$

alors

$$\frac{nd_G^2(\hat{\theta}_n(\cdot), \theta_0(\cdot)) - p}{\sqrt{2p}} \Rightarrow N(0,1).$$

2.5 Les instruments optimaux

Nous nous proposons dans cette partie d'estimer

$$B(x) = D'(x)\Omega^{-1}(x)$$

où

$$D(x) = \mathbb{E}\left[\frac{\partial g}{\partial \theta}\left(Y, \langle X, \theta_0 \rangle\right) | X = x\right]$$

est une matrice de taille $s \times p$ et où

$$\Omega(x) = \mathbb{E}\left[g(Y, \langle X, \theta_0 \rangle) g'(Y, \langle X, \theta_0 \rangle) | X = x\right]$$

est une matrice de taille $s \times s$. On suppose dans la suite que $s = 1$ pour alléger les notations.

Faisons une remarque sur les instruments optimaux dans le cas du modèle linéaire fonctionnel. Les instruments optimaux sont alors connus, en effet si on se place dans le cadre du modèle linéaire fonctionnel, $Y = \int_0^1 X(t)\theta_0(t)dt + e$, alors

$$g\left(Y, \int X\theta\right) = Y - \int X\theta,$$

ainsi

$$D(x) = x$$

et

$$\Omega(x) = \mathbb{E}(e^2|x) = \sigma^2(x),$$

où e est le terme d'erreur dans le modèle linéaire fonctionnel. Le fait de connaître les instruments optimaux dans le cas de modèle particulier renforce la partie d'estimation dans le cas d'instruments optimaux supposés connus. Il convient néanmoins d'estimer ces instruments dans le cas où ils ne sont pas connus. Nous supposons que les fonctions de régression à estimer ne dépendent que des p premières valeurs de la décomposition de X dans une base, voir ci-dessous la définition de la semi-métrique $d(\cdot, \cdot)$ et l'hypothèse (E)-b.

Soit Φ une fonction, on peut s'intéresser au cas, plus général, de l'estimation de

$$m(x, t) = \mathbb{E}[\Phi(Y, t)|X = x].$$

K est un noyau et h est un paramètre fenêtre associé à K. Nous pouvons maintenant définir l'estimateur :

$$\hat{m}(x, t) = \frac{\sum_{i=1}^{n} \Phi(Y_i, t) K(h^{-1}d(X_i, x))}{\sum_{i=1}^{n} K(h^{-1}d(X_i, x))}, \tag{2.5.6}$$

où x appartient à un sous-espace compact S de $L^2[0,1]$ pour la semi-norme induite par $d(\cdot,\cdot)$ et t est un réel. Nous établissons un résultat de convergence uniforme pour $\hat{m}$ vers m :

$$\sup_{t\in\mathbb{R},x\in S} |\hat{m}(x,t) - m(x,t)| = o_{\mathbb{P}}(1).$$

Ainsi, en disposant d'un estimateur initial $\hat{\theta}_n$ de θ_0, nous pouvons écrire

$$\begin{aligned}
& |\hat{m}(x,\langle x,\hat{\theta}_n\rangle) - m(x,\langle x,\theta_0\rangle)| \\
= \; & |\hat{m}(x,\langle x,\hat{\theta}_n\rangle) - \hat{m}(x,\langle x,\theta_0\rangle) + \hat{m}(x,\langle x,\theta_0\rangle) - m(x,\langle x,\theta_0\rangle)| \\
\leq \; & \{\sup_{t\in\mathbb{R},x\in S} |\hat{m}(x,t) - m(x,t)|\} + C\|x\|\|\hat{\theta}_n - \theta_0\| \\
\leq \; & \{\sup_{t\in\mathbb{R},x\in S} |\hat{m}(x,t) - m(x,t)|\} + C'\|\hat{\theta}_n - \theta_0\|,
\end{aligned}$$

où C et C' sont des constantes. La semi-distance $d(\cdot,\cdot)$ joue un rôle primordial. On choisit au préalable une base de $L^2[0,1]$ que l'on note (e_j) pour $j = 1,2,\ldots$. La semi-distance que l'on utilise est induite par la norme usuelle de $L^2[0,1]$. Rappelons que

$$\chi^{(k)} = \left(\chi_1 = \int_0^1 \chi(t)e_1(t)dt, \ldots, \chi_k = \int_0^1 \chi(t)e_k(t)dt\right)'.$$

Soit S un compact de $L^2[0,1]$ pour la distance d, définissons l'application ζ de $L^2[0,1]$ dans $\mathbb{R}^k$ qui à une fonction χ fait correspondre $\chi^{(k)}$, $S^{(k)} = \zeta(S)$ est alors un compact de $\mathbb{R}^k$ pour la norme 2. Pour tout ensemble I de $\mathbb{R}^k$, I^ε est l'ensemble des $\varepsilon-$voisins pour la norme max $|\cdot|_+$, qui est $|x|_+ = \max_{1\leq i\leq k}|x_i|$, pour $x \in \mathbb{R}^k$.

Faisons une remarque essentielle, on peut décomposer $\hat{m}(x,t) - m(x,t)$ de la façon suivante :

$$\hat{m}(x,t) - m(x,t) = \hat{m}(x,t) - \hat{m}(x^{(k)},t) + \hat{m}(x^{(k)},t) - m(x^{(k)},t) + m(x^{(k)},t) - m(x,t).$$

Ainsi, le terme $\hat{m}(x^{(k)},t) - m(x^{(k)},t)$ est ramené à un problème fini-dimensionnel que l'on traite à l'aide d'Einmahl et Mason (2005). Faisons maintenant quelques hypothèses.

Assumption E

(a) $\forall x \in S$, $0 < C\psi(h) \leq \mathbb{P}(X \in B(x,h)) \leq C'\psi(h) < \infty$, *pour* C, C' *des constantes.* ψ *vérifie*

$$\exists C^\star > 0, \exists \epsilon_0, \forall \epsilon < \epsilon_0, \quad \int_0^\epsilon \psi(u)du > C^\star \epsilon\psi(\epsilon),$$

$\psi(h) \to 0$ *quand* $h \to 0$ *et*

$$\frac{1}{h\psi(h)} \sup_{x\in S} \|x - x^{(k)}\| \to 0.$$

(b) $\forall x_1, x_2 \in S, \forall t \in \mathbb{R}$,

$$|m(x_1,t) - m(x_2,t)| \leq C(t)d(x_1,x_2)$$

avec $\sup_{t\in\mathbb{R}} C(t) < +\infty$.

(c) $\forall k$, $X^{(k)}$ *admet une densité continue et strictement positive sur* $S^{(k)\varepsilon}$, *pour un* $0 < \varepsilon < 1$.

(d) K *est positif, de support* $[0,1]$, *d'intégrale 1, sa dérivée existe et vérifie la condition* $-\infty < C \leq K'(t) \leq C' < 0$, *pour des constantes* C, C'. *De plus,* $nh^k \to +\infty$.

(e) L'enveloppe F *est défini par* $\sup_{t\in\mathbb{R}} |\Phi(y,t)| \leq F(y)$, $y \in \mathbb{R}$. *Elle vérifie*

$$\forall m \geq 2,\, \mathbb{E}\left[F^m(Y)|X = x\right] < \delta_m(x) < C < +\infty$$

avec δ_m *continue sur* S *et* C *une constante, de plus*

$$\sup_k \sup_{z\in S^{(k)\varepsilon}} \mathbb{E}(F^l(Y) \mid X^{(k)} = z) < \infty,$$

pour un $l > 2$.

Theorem 5.1 *Sous les hypothèses E, on obtient :*

$$\sup_{t\in\mathbb{R},x\in S} |\hat{m}(x,t) - m(x,t)| = o_{\mathbb{P}}(1).$$

Faisons quelques commentaires : la condition $nh^k \to +\infty$ est restrictive. h ne doit pas tendre vite vers 0 et k ne doit pas tendre rapidement vers l'infini. Par exemple, si on prend $h = 1/(\log n)^\tau$ alors $k << \log n/(\tau \log\log n)$. Comparons maintenant à l'article de Ferraty et al. (2010). Dans les cas usuels, leur condition (H5b) est vérifiée dés que $(\log n)^2 = O(n\psi(h))$. Cette dernière condition est moins restrictive que $nh^k \to +\infty$, car k tend vers l'infini. Cependant, leur condition (H2) est très restrictive. Car s'ils prennent une semi-métrique à partir de projections sur un espace de dimension k fixe (comme leur exemple 4), la condition (H2) est difficile à vérifier.

2.6 Simulations

La première étude que nous avons effectué consiste à estimer la fonction θ_0 par une méthode de Monte Carlo. On se fixe une base ϕ_j, $j \geq 1$ qui est une base de Fourier $\phi_j(t) = \sqrt{2}\sin(\pi j t)$, $t \in [0,1]$, $j \geq 1$. Nous définissons θ_0 dans cette base :

$$\theta_0(t) = \sum_{j=1}^{20} \theta_{0,j}\phi_j(t),$$

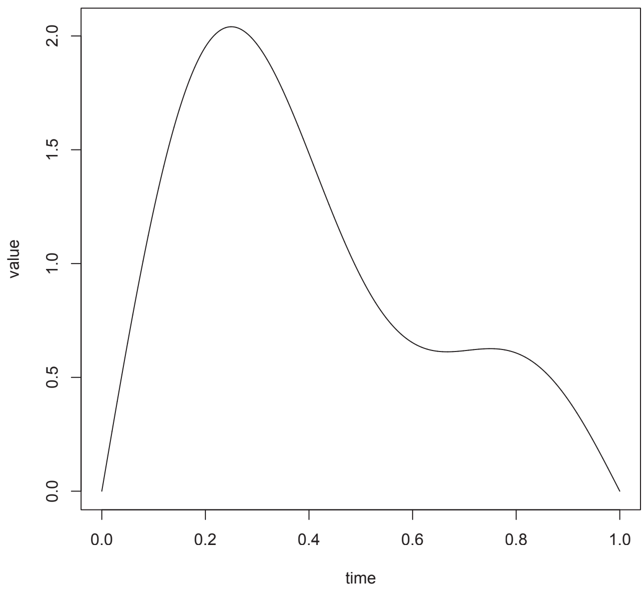

FIGURE 2.1 – La fonction θ_0.

où $\theta_{0,j} = 1/j$, $1 \leq j \leq 3$, $\theta_{0,j} = 0$ pour $j > 3$. La figure 2.1 représente la fonction θ_0. Nous construisons ensuite un échantillon (x_i, y_i) pour $i = 1, \ldots, 400$ comme suit

$$x(t) = \sum_{j=1}^{20} \epsilon_j \phi_j(t),$$

où les ϵ_j sont générés par une loi normale centrée et de variance $1/j^2$. On choisit un lien logit avec $\rho(t) = \exp(t)/(1+\exp(t))$ et les y_i sont générés par y qui suit une une loi de Bernoulli de paramètre $p(x) = \rho(\sum_{j=1}^{20} \theta_{0,j}\epsilon_j)$. Une fois obtenu l'échantillon, on estime θ_0 par la méthode décrite plus haut avec $p = 5$ et une base différente de ϕ_j à savoir une base spline à 5 éléments. La figure 2.2 représente les fonctions θ_0 et $\hat{\theta}$ obtenu en moyennant sur 1000 réplications la procédure.

2.7 Preuves

Dans toutes les preuves, on prend $\hat{P} = I_p$, sans restriction de généralité.

Preuve du lemme 4.1.

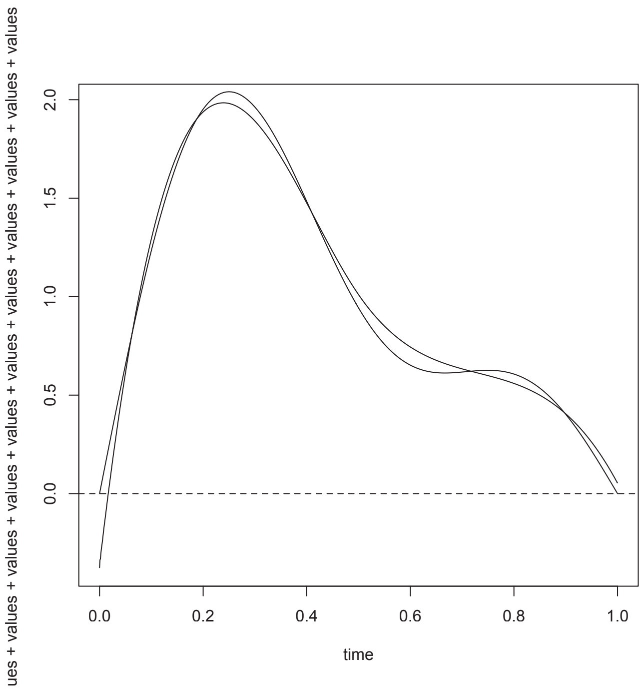

FIGURE 2.2 – Estimation de θ_0 avec les instruments optimaux.

On utilise une version du théorème 2.1 de Newey et McFadden (1994) :

Theorem 7.1 *On suppose qu'il existe une fonction aléatoire $\hat{Q}_n(\theta)$ tel que $\hat{\theta}$ maximise $\hat{Q}_n(\theta)$.On suppose de plus qu'il existe une fonction $Q_0(\theta)$ tel que :*

(i) θ_0 *est l'unique maximum de* $Q_0(\theta)$,

(ii) Θ *est compact,*

(iii) $Q_0(\theta)$ *est continue,*

(iv) on a

$$\sup_{\theta\in\Theta} |\hat{Q}_n(\theta) - Q_0(\theta)| \xrightarrow{\mathcal{P}} 0\,,$$

alors on obtient

$$\hat{\theta} \xrightarrow{\mathcal{P}} \theta_0\,.$$

Pour obtenir la convergence uniforme en probabilité de $\hat{Q}_n(\theta)$, on a besoin d'une loi uniforme des grands nombres, voir le lemme 2.4 de Newey et McFadden (1994) :

Lemma 7.2 *Sous les hypothèses suivantes :*

1. *les données sont i.i.d.,*

2. *Θ est compact,*
3. *$a(z_i, \theta)$ est continue en chaque $\theta \in \Theta$ avec probabilité un,*
4. *Il existe $d(z)$ avec $\|a(z,\theta)\| \leq d(z)$ $\forall \theta \in \Theta$ et $\mathbb{E}[d(z)] < \infty$,*

on obtient la continuité de $\mathbb{E}[a(z,\theta)]$ et

$$\sup_{\theta \in \Theta} \| \frac{1}{n} \sum_{i=1}^{n} a(z_i, \theta) - \mathbb{E}[a(z,\theta)] \| \xrightarrow{\mathcal{P}} 0 \,.$$

On pose $\hat{Q}_n(\theta) = \hat{g}'_n(\theta^{(p)}) \hat{g}_n(\theta^{(p)})$, avec $\hat{g}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} B(x_i^{(p)}) g(y_i, \langle x_i^{(p)}, \theta^{(p)} \rangle)$. Posons de plus

$$Q_0(\theta) = \lim_{n \to \infty} \| \mathbb{E} \left[B(x^{(p)}) g(y, \langle x^{(p)}, \theta^{(p)} \rangle) \right] \|^2,$$

on a $Q_0(\theta_0) = 0$ qui est l'unique minimum. En effet, s'il existe θ_1 tel que $Q_0(\theta_1) = 0$, alors en notant B_j la j-ème coordonnée de B, nous avons

$$\mathbb{E} \left[B_j(x^{(p)}) \mathbb{E}[g(y, \langle x^{(p)}, \theta_1^{(p)} \rangle \mid x^{(p)}] \right] \to 0.$$

D'où $\forall j$,

$$\mathbb{E} \left[B_j(x) \mathbb{E}[g(y, \langle x, \theta_1 \rangle \mid x] \right] = 0.$$

D'où $\mathbb{E}[g(y, \langle x, \theta_1 \rangle \mid x] = 0$ et donc $\theta_1 = \theta_0$.
On applique ensuite la loi uniforme des grands nombres, on a alors par l'hypothèse 4 qui est vérifiée par la condition (d) du lemme 4.1 :

$$\sup_{\theta \in \Theta} \| \hat{g}_n(\theta^{(p)}) - g_p(\theta^{(p)}) \| \xrightarrow{\mathcal{P}} 0 \,,$$

où $g_p(\theta^{(p)}) = \mathbb{E}[B(x^{(p)}) g(y, \langle x^{(p)}, \theta^{(p)} \rangle)]$. On a

$$\begin{aligned} |\hat{Q}_n(\theta) - Q_0(\theta)| &\leq |\hat{Q}_n(\theta) - g'_p(\theta^{(p)}) g_p(\theta^{(p)})| + |g'_p(\theta^{(p)}) g_p(\theta^{(p)}) - Q_0(\theta)| \\ &\leq \|\hat{g}_n(\theta^{(p)}) - g_p(\theta^{(p)})\|^2 + 2\|g_p(\theta^{(p)})\| \|\hat{g}_n(\theta^{(p)}) - g_p(\theta^{(p)})\| \\ &\quad + |g'_p(\theta^{(p)}) g_p(\theta^{(p)}) - Q_0(\theta)|. \end{aligned}$$

D'où

$$\sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - Q_0(\theta)| \xrightarrow[n\infty]{\mathcal{P}} 0.$$

Toutes les hypothèses du théorème sont vérifiées, on en déduit que

$$\|\hat{\theta}_n - \theta_0^{(p)}\| \xrightarrow[n\infty]{\mathcal{P}} 0.$$

Preuve du théorème 4.2.

En posant

$$U(\theta)=\frac{\partial \hat{g}_n}{\partial \theta'}(\hat{\theta})\hat{P}\hat{g}_n(\theta),$$

on a par Taylor, pour un $\bar{\theta}$ compris entre $\tilde{\theta}$ et $\hat{\theta}$

$$0=U(\hat{\theta})=U(\theta_0)+\frac{\partial U}{\partial \theta}(\bar{\theta})\left(\hat{\theta}-\theta_0\right).$$

D'où

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right)=-\left(\frac{\partial \hat{g}_n}{\partial \theta'}(\hat{\theta})\hat{P}\frac{\partial \hat{g}_n}{\partial \theta}(\bar{\theta})\right)^{-1}\sqrt{n}U(\theta_0),$$

qui se simplifie en

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right)=-\left(\frac{\partial \hat{g}_n}{\partial \theta}(\bar{\theta})\right)^{-1}\sqrt{n}\hat{g}_n(\theta_0).$$

Dans une première partie, on s'attache à remplacer $\bar{\theta}$ par θ_0. $\sqrt{n}\left(\hat{\theta}-\theta_0\right)$ s'écrit

$$=-\left[\frac{\partial \hat{g}_n}{\partial \theta}(\bar{\theta})-\frac{\partial \hat{g}_n}{\partial \theta}(\theta_0)+\frac{\partial \hat{g}_n}{\partial \theta}(\theta_0)\right]^{-1}\sqrt{n}\hat{g}_n(\theta_0),$$

$$=-\left\{I_p+\left[\frac{\partial \hat{g}_n}{\partial \theta}(\theta_0)\right]^{-1}\left[\frac{\partial \hat{g}_n}{\partial \theta}(\bar{\theta})-\frac{\partial \hat{g}_n}{\partial \theta}(\theta_0)\right]\right\}^{-1}\left[\frac{\partial \hat{g}_n}{\partial \theta}(\theta_0)\right]^{-1}\sqrt{n}\hat{g}_n(\theta_0).$$

Comme pour toute matrice C avec $\|C\|_2<1$ on a $(I_p+C)^{-1}=I_p-C+R$ avec $\|R\|_2\leq(1-\|C\|_2)^{-1}\|C\|_2^2$, il suffit de montrer que

$$\left\|\left[\frac{\partial \hat{g}_n}{\partial \theta}(\theta_0)\right]^{-1}\right\|_2=O_p(1)\ \text{ and }\ \left\|\frac{\partial \hat{g}_n}{\partial \theta}(\bar{\theta})-\frac{\partial \hat{g}_n}{\partial \theta}(\theta_0)\right\|_2=o_p(1). \tag{2.7.7}$$

Par la loi de grands nombres,

$$\frac{\partial \hat{g}_n}{\partial \theta}(\theta_0)-\mathbb{E}\left[\frac{\partial \hat{g}_n}{\partial \theta}(\theta_0)\right]=\frac{\partial \hat{g}_n}{\partial \theta}(\theta_0)-\Lambda^{-1}=o_p(1).$$

Ceci assure la première partie de l'équation (2.7.7). Pour la deuxième partie de l'équation (2.7.7), on peut écrire

$$\begin{aligned}
\left\|\frac{\partial\hat{g}_n}{\partial\theta}(\bar{\theta})-\frac{\partial\hat{g}_n}{\partial\theta}(\theta_0)\right\|_2 &= \frac{1}{n}\left\|\sum_{i=1}^{n}B(x_i)(\frac{\partial g}{\partial\theta}(z_i,\bar{\theta})-\frac{\partial g}{\partial\theta}(z_i,\theta_0))\right\|_2\\
&\leq \frac{1}{n}\sum_{i=1}^{n}\|B(x_i)\|\left\|\frac{\partial g}{\partial\theta}(z_i,\bar{\theta})-\frac{\partial g}{\partial\theta}(z_i,\theta_0)\right\|_2\\
&\leq O_p(\|\bar{\theta}-\theta_0\|^a)\frac{1}{n}\sum_{i=1}^{n}\|B(x_i)\||c(y_i)|\\
&= o_p(1),
\end{aligned}$$

avec les hypothèses

$$\left\|\frac{\partial g}{\partial\theta}(z_i,b)-\frac{\partial g}{\partial\theta}(z_i,d)\right\|_2\leq c(y_i)\|b-d\|^a$$

et

$$\|B_j(x_i)\||c(y_i)|\leq\Gamma(z_i),\qquad \mathbb{E}[\Gamma(z_i)]<\infty,$$

car

$$\left\|\frac{\partial g}{\partial\theta}(z_i,\sum_{j=1}^{p}\epsilon_j^i\bar{\theta}_j)-\frac{\partial g}{\partial\theta}(z_i,\sum_{j=1}^{p}\epsilon_j^i\theta_{0,j})\right\|_2^2\leq c(y_i)|\sum_{j=1}^{p}\epsilon_j^i(\bar{\theta}_j-\theta_{0,j})|^a$$

or par Cauchy-Schwartz, on a

$$|\sum_{j=1}^{p}\epsilon_j^i(\bar{\theta}_j-\theta_{0,j})|\leq\sqrt{\sum_{j=1}^{+\infty}\epsilon_j^{i2}}\sqrt{\sum_{j=1}^{p}(\bar{\theta}_j-\theta_{0,j})^2}=O_p(1).$$

On étudie maintenant $Z_n^{'}\Lambda^{-1}Z_n$ où $Z_n=(\frac{\partial\hat{g}_n}{\partial\theta}(\theta_0))^{-1}\sqrt{n}\hat{g}_n(\theta_0)$. On cherche quel terme est dominant dans la décomposition suivante de $Z_n^{'}\Lambda^{-1}Z_n$. En posant $\chi_n=\Lambda^{\frac{1}{2}}\sqrt{n}\hat{g}_n(\theta_0)$ et $\psi_n=\Lambda^{-\frac{1}{2}}(\frac{\partial\hat{g}_n}{\partial\theta}(\theta_0))^{-1}\Lambda^{-\frac{1}{2}}$, on a alors la décomposition :

$$Z_n^{'}\Lambda^{-1}Z_n=\chi_n^{'}\chi_n+2\chi_n^{'}(\psi_n-I_p)\chi_n+\chi_n^{'}(\psi_n-I_p)(\psi_n-I_p)\chi_n.$$

Il reste à montrer que $||\psi_n-I_p||_2^2=O_p(\frac{1}{p})$.

Étudions d'abord ψ_n^{-1} :

$$\begin{aligned}
\psi_n^{-1} &= \Lambda^{-1/2}\frac{\partial\hat{g}_n}{\partial\theta}(\theta_0)\Lambda^{-1/2}\\
&= \left(\sum_{k_1,k_2=1}^{p}\xi_{l_1,k_1}^{(1/2)}(1/n\sum_{k=1}^{n}(\frac{\partial\tilde{g}_k}{\partial\theta})_{k_1,k_2})\xi_{k_2,l_2}^{(1/2)}\right)_{l_1,l_2=1}^{p}\\
&= \left(\frac{1}{n}\sum_{k_1,k_2=1}^{p}\xi_{l_1,k_1}^{(1/2)}\xi_{k_2,l_2}^{(1/2)}\sum_{k=1}^{n}(\frac{\partial\tilde{g}_k}{\partial\theta})_{k_1,k_2}\right)_{l_1,l_2=1}^{p}
\end{aligned}$$

d'où, le calcul suivant :

$$
\begin{aligned}
\mathbb{E}(||\psi_n^{-1} - I_p||_2^2) &= \mathbb{E}(\sum_{l_1,l_2=1}^{p} (\frac{1}{n}\sum_{k=1}^{n}\sum_{k_1,k_2=1}^{p} \xi_{l_1,k_1}^{(1/2)}\xi_{k_2,l_2}^{(1/2)}(\frac{\partial \tilde{g}_k}{\partial \theta})_{k_1,k_2} - \delta_{l_1,l_2}) \\
&\quad \times(\frac{1}{n}\sum_{k'=1}^{n}\sum_{k_1',k_2'=1}^{p} \xi_{l_1,k_1'}^{(1/2)}\xi_{k_2',l_2}^{(1/2)}(\frac{\partial \tilde{g}_k'}{\partial \theta})_{k_1',k_2'} - \delta_{l_1,l_2}) \\
&= \mathbb{E}(\sum_{l_1,l_2=1}^{p} (\frac{1}{n^2}\sum_{k,k'=1}^{n}\sum_{k_1,k_2,k_1',k_2'}^{p} \xi_{l_1,k_1}^{(1/2)}\xi_{k_2,l_2}^{(1/2)}\xi_{l_1,k_1'}^{(1/2)}\xi_{k_2',l_2}^{(1/2)}(\frac{\partial \tilde{g}_k}{\partial \theta})_{k_1,k_2}(\frac{\partial \tilde{g}_k'}{\partial \theta})_{k_1',k_2'} \\
&\quad -2\delta_{l_1,l_2}\frac{1}{n}\sum_{k=1}^{n}\sum_{k_1,k_2=1}^{p} \xi_{l_1,k_1}^{(1/2)}\xi_{k_2,l_2}^{(1/2)}(\frac{\partial \tilde{g}_k}{\partial \theta})_{k_1,k_2} \\
&\quad +\delta_{l_1,l_2}) \\
&= \sum_{l_1,l_2=1}^{p} (\frac{1}{n^2}\sum_{k,k'=1}^{n}\sum_{k_1,k_2,k_1',k_2'}^{p} \xi_{l_1,k_1}^{(1/2)}\xi_{k_2,l_2}^{(1/2)}\xi_{l_1,k_1'}^{(1/2)}\xi_{k_2',l_2}^{(1/2)}\mathbb{E}[(\frac{\partial \tilde{g}_k}{\partial \theta})_{k_1,k_2}(\frac{\partial \tilde{g}_k'}{\partial \theta})_{k_1',k_2'}] \\
&\quad -2\delta_{l_1,l_2}\frac{1}{n}\sum_{k=1}^{n}\sum_{k_1,k_2=1}^{p} \xi_{l_1,k_1}^{(1/2)}\xi_{k_2,l_2}^{(1/2)}\Lambda_{k_1,k_2}^{-1} \\
&\quad +\delta_{l_1,l_2})
\end{aligned}
$$

car $\mathbb{E}[(\frac{\partial \tilde{g}_k}{\partial \theta})_{k_1,k_2}] = \Lambda_{k_1,k_2}^{-1}$
ainsi,

$$
-2\delta_{l_1,l_2}\frac{1}{n}\sum_{k=1}^{n}\sum_{k_1,k_2=1}^{p} \xi_{l_1,k_1}^{(1/2)}\xi_{k_2,l_2}^{(1/2)}\Lambda_{k_1,k_2}^{-1} = -2\delta_{l_1,l_2}\sum_{k_1,k_2=1}^{p} \xi_{l_1,k_1}^{(1/2)}\xi_{k_2,l_2}^{(1/2)}\Lambda_{k_1,k_2}^{-1} = -2\delta_{l_1,l_2}^2 = -2\delta_{l_1,l_2}
$$

d'où en reprenant le calcul de $\mathbb{E}(||\psi_n^{-1} - I_p||_2^2)$, et en scindant en deux termes suivant $k = k'$ ou $k \neq k'$, c-à-d suivant que les variables sont indépendantes ou non, on obtient :

$$
\begin{aligned}
& \mathbb{E}(||\psi_n^{-1} - I_p||_2^2) \\
= & \sum_{l_1,l_2=1}^{p} (\frac{1}{n^2} \sum_{k=1}^{n} \sum_{k_1,k_2,k_1',k_2'}^{p} \xi_{l_1,k_1}^{(1/2)} \xi_{k_2,l_2}^{(1/2)} \xi_{l_1,k_1'}^{(1/2)} \xi_{k_2',l_2}^{(1/2)} \mathbb{E}[(\frac{\partial \tilde{g}_k}{\partial \theta})_{k_1,k_2} (\frac{\partial \tilde{g}_k}{\partial \theta})_{k_1',k_2'}] \\
+ & \frac{1}{n^2} \sum_{k \neq k'}^{n} \sum_{k_1,k_2,k_1',k_2'}^{p} \xi_{l_1,k_1}^{(1/2)} \xi_{k_2,l_2}^{(1/2)} \xi_{l_1,k_1'}^{(1/2)} \xi_{k_2',l_2}^{(1/2)} \Lambda_{k_1,k_2}^{-1} \Lambda_{k_1',k_2'}^{-1} - \delta_{l_1,l_2}) \\
= & \sum_{l_1,l_2=1}^{p} \frac{1}{n^2} \sum_{k=1}^{n} \sum_{k_1,k_2,k_1',k_2'}^{p} \xi_{l_1,k_1}^{(1/2)} \xi_{k_2,l_2}^{(1/2)} \xi_{l_1,k_1'}^{(1/2)} \xi_{k_2',l_2}^{(1/2)} \mathbb{E}[(\frac{\partial \tilde{g}_k}{\partial \theta})_{k_1,k_2} (\frac{\partial \tilde{g}_k}{\partial \theta})_{k_1',k_2'}] - \frac{1}{n} \delta_{l_1,l_2} \\
= & \frac{1}{n^2} \sum_{l_1,l_2=1}^{p} \sum_{k=1}^{n} \sum_{k_1,k_2,k_1',k_2'}^{p} \xi_{l_1,k_1}^{(1/2)} \xi_{k_2,l_2}^{(1/2)} \xi_{l_1,k_1'}^{(1/2)} \xi_{k_2',l_2}^{(1/2)} \mathbb{E}[(\frac{\partial \tilde{g}_k}{\partial \theta})_{k_1,k_2} (\frac{\partial \tilde{g}_k}{\partial \theta})_{k_1',k_2'}] + O(\frac{p}{n}) \\
= & \frac{1}{n} \sum_{k_1,k_2,k_1',k_2'}^{p} \xi_{k_1,k_1'} \xi_{k_2,k_2'} \mathbb{E}[(\frac{\partial \tilde{g}_k}{\partial \theta})_{k_1,k_2} (\frac{\partial \tilde{g}_k}{\partial \theta})_{k_1',k_2'}] + O(\frac{p}{n})
\end{aligned}
$$

par hypothèses $\sum_{k_1,k_2,k_1',k_2'}^{p} \xi_{k_1,k_1'} \xi_{k_2,k_2'} \mathbb{E}[(\frac{\partial \tilde{g}_k}{\partial \theta})_{k_1,k_2} (\frac{\partial \tilde{g}_k}{\partial \theta})_{k_1',k_2'}] = o(n/p^2)$. On a donc le résultat suivant $||\psi_n^{-1} - I_p||_2^2 = O_p(1/p^2)$.On utilise ensuite le résultat suivant :

$$||\psi_n - I_p||_2 \leq ||\psi_n||_2 ||\psi_n^{-1} - I_p||_2$$

conjugué avec ce résultat

$$||\psi_n||_2 \leq \frac{||I_p||_2}{1 - ||\psi_n^{-1} - I_p||_2},$$

cela nous donne le résultat escompté, à savoir $||\psi_n - I_p||_2 = O_p(\frac{1}{p})$.

Traitons le terme dominant : On va montrer que

$$\frac{\chi_n' \chi_n - p}{\sqrt{2p}} \longrightarrow N(0,1).$$

Pour se faire, on décompose en deux termes :

$$
\begin{aligned}
\chi_n' \chi_n &= n \hat{g}_n(\theta_0)' \Lambda \hat{g}_n(\theta_0) \\
&= \frac{1}{n} \sum_{i=1}^{n} (B(x_i) g(z_i, \theta_0))' \Lambda (B(x_i) g(z_i, \theta_0)) \\
&+ \frac{1}{n} \sum_{i \neq k}^{n} (B(x_i) g(z_i, \theta_0))' \Lambda (B(x_k) g(z_k, \theta_0)) \\
&= A_n + B_n
\end{aligned}
$$

on a $\mathbb{E}[A_n] = p$ ainsi que $Var(A_n) = o(p)$ ce qui nous permet d'écrire

$$\frac{A_n - p}{\sqrt{p}} \longrightarrow 0$$

En effet,

$$\begin{aligned}
\mathbb{E}[A_n] &= \mathbb{E}[g'(z_1, \theta_0)B'(x_1)\Lambda B(x_1)g(z_1, \theta_0)] \\
&= \mathbb{E}[tr(g'B'\Lambda Bg)] \\
&= \mathbb{E}[tr(g'B'\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}Bg)] \\
&= \mathbb{E}[tr(\Lambda^{\frac{1}{2}}Bgg'B'\Lambda^{\frac{1}{2}})] \\
&= tr(\Lambda^{\frac{1}{2}}\mathbb{E}[Bgg'B']\Lambda^{\frac{1}{2}})
\end{aligned}$$

ensuite on a $\mathbb{E}[Bgg'B'] = \mathbb{E}[\mathbb{E}[Bgg'B']|x_1]$ et donc $\mathbb{E}[Bgg'B'] = \Lambda^{-1}$

$$\begin{aligned}
\mathbb{E}[A_n] &= tr(\Lambda^{\frac{1}{2}}\Lambda^{-1}\Lambda^{\frac{1}{2}}) \\
&= tr(I_p) \\
&= p
\end{aligned}$$

On calcule maintenant la variance, pour cela calculons $\mathbb{E}[A_n^2]$

$$A_n^2 = \frac{1}{n^2}\sum_{i,j=1}^{n} \tilde{g}'^{(i)}\Lambda\tilde{g}^{(i)}\tilde{g}'^{(j)}\Lambda\tilde{g}^{(j)} \tag{2.7.8}$$

$$\begin{aligned}
\mathbb{E}[A_n^2] &= \frac{1}{n^2}\sum_{i,j=1}^{n} \mathbb{E}(\tilde{g}'^{(i)}\Lambda\tilde{g}^{(i)}\tilde{g}'^{(j)}\Lambda\tilde{g}^{(j)}) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \mathbb{E}(\tilde{g}'^{(i)}\Lambda\tilde{g}^{(i)}\tilde{g}'^{(i)}\Lambda\tilde{g}^{(i)}) \\
&+ \frac{1}{n^2}\sum_{i\neq j}^{n} \mathbb{E}(\tilde{g}'^{(i)}\Lambda\tilde{g}^{(i)}\tilde{g}'^{(j)}\Lambda\tilde{g}^{(j)})
\end{aligned}$$

Or $\mathbb{E}(\tilde{g}'^{(i)}\Lambda\tilde{g}^{(i)}\tilde{g}'^{(j)}\Lambda\tilde{g}^{(j)}) = (\mathbb{E}(\tilde{g}'^{(i)}\Lambda\tilde{g}^{(i)}))^2 = p^2$

$$\begin{aligned}
\mathbb{E}[A_n^2] &= \frac{n(n-1)}{n^2}p^2 + \frac{1}{n^2}\sum_{i=1}^{n} \mathbb{E}(\tilde{g}'^{(i)}\Lambda\tilde{g}^{(i)}\tilde{g}'^{(i)}\Lambda\tilde{g}^{(i)}) \\
&= p^2 - \frac{p^2}{n} + \frac{1}{n}\mathbb{E}(\tilde{g}'^{(1)}\Lambda\tilde{g}^{(1)}\tilde{g}'^{(1)}\Lambda\tilde{g}^{(1)})
\end{aligned}$$

On utilise alors l'hypothèse D-(e), en effet

$$\mathbb{E}(\tilde{g}'^{(1)}\Lambda\tilde{g}^{(1)}\tilde{g}'^{(1)}\Lambda\tilde{g}^{(1)}) = \sum_{k_1,k_2,k_3,k_4=1}^{p} \mathbb{E}[\tilde{g}_{k_1}\tilde{g}_{k_2}\tilde{g}_{k_3}\tilde{g}_{k_4}]\xi_{k_1,k_2}\xi_{k_3,k_4} = o(n/p^2) \quad (2.7.9)$$

On obtient donc que $\mathrm{Var}(A_n) = o(p)$ d'où le résultat.

En utilisant un T.C.L pour martingales démontré par Brown (1971), on montre que $B_n/\sqrt{2p} \to N(0,1)$. Rappelons ce que vaut B_n :

$$B_n = \frac{1}{n}\sum_{i\neq k}^{n}(B(x_i)g(z_i,\theta_0))'\Lambda(B(x_k)g(z_k,\theta_0)),$$

Il faut centrer pour pouvoir appliquer le théorème. Pour cela, posons

$$\widetilde{B_n} = \frac{1}{n}\sum_{i\neq k}\tilde{g}_c'^{(i)}\Lambda\tilde{g}_c^{(k)},$$

où $\tilde{g}_c^{(j)} = B(x_j)g(z_j,\theta_0) - \mathbb{E}[B(x_1)g(z_1,\theta_0)]$. On obtient alors la décomposition de B_n en quatre termes :

$$\begin{aligned} B_n = \widetilde{B_n} \quad &+ \quad \frac{n-1}{n}\left(\mathbb{E}B(x_1)g(z_1,\theta_0)\right)'\Lambda\left(\sum_{k=1}^{n}B(x_k)g(z_k,\theta_0)\right) \\ &+ \quad \frac{n-1}{n}\left(\sum_{k=1}^{n}B(x_k)g(z_k,\theta_0)\right)'\Lambda\left(\mathbb{E}B(x_1)g(z_1,\theta_0)\right) \\ &- \quad (n-1)\left(\mathbb{E}B(x_1)g(z_1,\theta_0)\right)'\Lambda\left(\mathbb{E}B(x_1)g(z_1,\theta_0)\right) \end{aligned}$$

Remarquons que :

$$\mathbb{E}[B(x_1)g(z_1,\theta_0)] = O(pr_n).$$

Ainsi le dernier terme est de l'ordre $O(np_n^2r_n^2)$, le second et le troisième terme sont de l'ordre $O_{\mathbb{P}}(np_n^2r_n^2)$. D'où en divisant par $\sqrt{2p}$ et en utilisant l'hypothèse $n^{11/8}r_n^2 \to 0$, ces termes sont négligeables par rapport à $\widetilde{B_n}$. Etudions maintenant $\widetilde{B_n}$. Posons

$$W_{nj} = \sum_{k=1}^{j-1}\tilde{g}_c^{(j)\prime}\Lambda\tilde{g}_c^{(k)}$$

on obtient alors $\widetilde{B_n} = \frac{2}{n}\sum_{j=1}^{n}W_{nj}$.
Les $\{W_{nj}, 1 \leq j \leq n, n \in \mathbb{N}\}$ forment un tableau triangulaire de différences de

martingales associée aux filtrations $\mathcal{F}_{nj} = \sigma(\epsilon_t^{(i)}, y_i, 1 \leq i \leq j, 1 \leq t \leq p)(1 \leq j \leq n, n \in \mathbb{N})$.
En effet, $W_{nj} \in \mathcal{F}_{nj}$ et

$$\mathbb{E}[W_{nj}|\mathcal{F}_{n,j-1}] = 0,$$

car $\mathbb{E}\left[\tilde{g}_c^{(j)}|\mathcal{F}_{n,j-1}\right] = 0$. Pour appliquer le théorème de Brown (1971), il faut valider la condition de normalisation conditionnelle ainsi que la condition de Lyapunov conditionnelle. En écrivant, $\widetilde{W_{nj}} = \frac{2}{n\sqrt{2p}} W_{nj}$, la condition de normalisation conditionnelle s'écrit :

$$\sum_{j=1}^{n} \mathbb{E}[\widetilde{W_{nj}}^2|\mathcal{F}_{n,j-1}] \xrightarrow{p} 1,$$

la condition de Lyapunov conditionnelle s'écrit :

$$\sum_{j=1}^{n} \mathbb{E}[\widetilde{W_{nj}}^4|\mathcal{F}_{n,j-1}] \xrightarrow{p} 0.$$

Pour cela calculons W_{nj}^2 :

$$\begin{aligned}
W_{nj}^2 &= \sum_{k,k'=1}^{j-1} \tilde{g}_c^{\prime(j)} \Lambda \tilde{g}_c^{(k)} \tilde{g}_c^{\prime(j)} \Lambda \tilde{g}_c^{(k')} \\
&= \sum_{k,k'=1}^{j-1} \sum_{t_1,t_2,t_3,t_4}^{p} \tilde{g}_{c,t_1}^{(j)} \xi_{t_1,t_2} \tilde{g}_{c,t_2}^{(k)} \tilde{g}_{c,t_3}^{(j)} \xi_{t_3,t_4} \tilde{g}_{c,t_4}^{(k')} \\
&= \sum_{k,k'=1}^{j-1} \sum_{t_1,t_2,t_3,t_4}^{p} \tilde{g}_{c,t_1}^{(j)} \tilde{g}_{c,t_3}^{(j)} \tilde{g}_{c,t_2}^{(k)} \tilde{g}_{c,t_4}^{(k')} \xi_{t_1,t_2} \xi_{t_3,t_4},
\end{aligned}$$

d'où, en prenant l'espérance conditionnelle,

$$\mathbb{E}[W_{nj}^2|\mathcal{F}_{n,j-1}] = \sum_{k,k'=1}^{j-1} \sum_{t_1,t_2,t_3,t_4}^{p} \mathbb{E}[\tilde{g}_{c,t_1}^{(j)} \tilde{g}_{c,t_3}^{(j)}] \tilde{g}_{c,t_2}^{(k)} \tilde{g}_{c,t_4}^{(k')} \xi_{t_1,t_2} \xi_{t_3,t_4}.$$

Or,

$$\mathbb{E}[\tilde{g}_{c,t_1}^{(j)} \tilde{g}_{c,t_3}^{(j)}] = (\Lambda^{-1})_{t_1,t_3} - \mathbb{E}[\tilde{g}_{t_1}^{(j)}]\mathbb{E}[\tilde{g}_{t_3}^{(j)}],$$

donc :

$$\begin{aligned}
\mathbb{E}[W_{nj}^2|\mathcal{F}_{n,j-1}] &= \sum_{k,k'=1}^{j-1} \sum_{t_2,t_4}^{p} \tilde{g}_{c,t_2}^{(k)} \tilde{g}_{c,t_4}^{(k')} \xi_{t_2,t_4} \\
&- \sum_{k,k'=1}^{j-1} \sum_{t_1,t_2,t_3,t_4}^{p} \mathbb{E}[\tilde{g}_{t_1}^{(j)}]\mathbb{E}[\tilde{g}_{t_3}^{(j)}] \tilde{g}_{c,t_2}^{(k)} \tilde{g}_{c,t_4}^{(k')} \xi_{t_1,t_2} \xi_{t_3,t_4}
\end{aligned}$$

On peut donc calculer son espérance :

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[W_{nj}^2|\mathcal{F}_{n,j-1}]] &= \sum_{k=1}^{j-1}\sum_{t_2,t_4}^{p}\mathbb{E}[\tilde{g}_{c,t_2}^{(k)}\tilde{g}_{c,t_4}^{(k)}]\xi_{t_2,t_4} \\
&- \sum_{k=1}^{j-1}\sum_{t_1,t_2,t_3,t_4}^{p}\mathbb{E}[\tilde{g}_{t_1}^{(j)}]\mathbb{E}[\tilde{g}_{t_3}^{(j)}]\mathbb{E}[\tilde{g}_{c,t_2}^{(k)}\tilde{g}_{c,t_4}^{(k')}]\xi_{t_1,t_2}\xi_{t_3,t_4} \\
&= (j-1)p-(j-1)\sum_{t_2,t_4}^{p}\mathbb{E}[\tilde{g}_{t_2}^{(k)}]\mathbb{E}[\tilde{g}_{t_4}^{(k)}]\xi_{t_2,t_4} \\
&- (j-1)\sum_{t_1,t_2,t_3,t_4}^{p}\mathbb{E}[\tilde{g}_{t_1}^{(1)}]\mathbb{E}[\tilde{g}_{t_3}^{(1)}]\mathbb{E}[\tilde{g}_{c,t_2}^{(1)}\tilde{g}_{c,t_4}^{(1)}]\xi_{t_1,t_2}\xi_{t_3,t_4}
\end{aligned}$$

maintenant,

$$\begin{aligned}
& \mathbb{E}[\sum_{j=1}^{n}\mathbb{E}[\widetilde{W_{nj}}^2|\mathcal{F}_{n,j-1}]] \\
= \quad & \sum_{j=1}^{n}\mathbb{E}[\frac{2}{n^2p}W_{nj}^2|\mathcal{F}_{n,j-1}]] \\
= \quad & \frac{2}{n^2p}\sum_{j=1}^{n}(j-1) \\
& \times\left(p-\sum_{t_2,t_4}^{p}\mathbb{E}[\tilde{g}_{t_2}^{(k)}]\mathbb{E}[\tilde{g}_{t_4}^{(k)}]\xi_{t_2,t_4}-\sum_{t_1,t_2,t_3,t_4}^{p}\mathbb{E}[\tilde{g}_{t_1}^{(1)}]\mathbb{E}[\tilde{g}_{t_3}^{(1)}]\mathbb{E}[\tilde{g}_{c,t_2}^{(1)}\tilde{g}_{c,t_4}^{(1)}]\xi_{t_1,t_2}\xi_{t_3,t_4}\right) \\
= \quad & \frac{n(n-1)}{n^2}\left(1-\frac{1}{p}O(p_n^2r_n^2)-\frac{1}{p}O(p_n^4r_n^4)\right) \\
\longrightarrow \quad & 1
\end{aligned}$$

Nous pouvons maintenant calculer la variance. Pour cela calculons, pour $k<j$:

$$\mathbb{E}[\mathbb{E}[W_{nj}^2|\mathcal{F}_{n,j-1}]\mathbb{E}[W_{nk}^2|\mathcal{F}_{n,k-1}].$$

En notant

$$\begin{aligned}
\mathbb{E}[W_{nj}^2|\mathcal{F}_{n,j-1}] &= \sum_{k,k'=1}^{j-1}\sum_{t_2,t_4}^{p}\tilde{g}_{c,t_2}^{(k)}\tilde{g}_{c,t_4}^{(k')}\xi_{t_2,t_4} \\
&- \sum_{k,k'=1}^{j-1}\sum_{t_1,t_2,t_3,t_4}^{p}\mathbb{E}[\tilde{g}_{t_1}^{(j)}]\mathbb{E}[\tilde{g}_{t_3}^{(j)}]\tilde{g}_{c,t_2}^{(k)}\tilde{g}_{c,t_4}^{(k')}\xi_{t_1,t_2}\xi_{t_3,t_4} \\
&= M_j-N_j,
\end{aligned}$$

on peut écrire

$$\begin{aligned}\mathbb{E}[\mathbb{E}[W_{nj}^2|\mathcal{F}_{n,j-1}]\mathbb{E}[W_{nk}^2|\mathcal{F}_{n,k-1}] &= \mathbb{E}[(M_j - N_j)(M_k - N_k))] \\ &= \mathbb{E}[M_jM_k - M_jN_k - M_kN_j + N_jN_k],\end{aligned}$$

où le terme dominant de cette somme est $\mathbb{E}[M_jM_k]$. Les trois autres termes sont négligeables par la condition D-(b). On calcule $\mathbb{E}[M_jM_k]$:

$$\begin{aligned}\mathbb{E}[M_jM_k] &= \sum_{i_1,i_2}^{j-1}\sum_{i_3,i_4}^{k-1}\sum_{t_1,t_2,t_3,t_4}^{p} \mathbb{E}[\tilde{g}_{c,t_1}^{(i_1)}\tilde{g}_{c,t_2}^{(i_2)}\tilde{g}_{c,t_3}^{(i_3)}\tilde{g}_{c,t_4}^{(i_4)}]\xi_{t_1,t_2}\xi_{t_3,t_4} \\ &= (k-1)\sum_{t_1,t_2,t_3,t_4}^{p} \mathbb{E}[\tilde{g}_{c,t_1}\tilde{g}_{c,t_2}\tilde{g}_{c,t_3}\tilde{g}_{c,t_4}]\xi_{t_1,t_2}\xi_{t_3,t_4} \\ &+ (j-1)(k-1)\sum_{t_1,t_2,t_3,t_4}^{p} \mathbb{E}[\tilde{g}_{c,t_1}\tilde{g}_{c,t_2}]\mathbb{E}[\tilde{g}_{c,t_3}\tilde{g}_{c,t_4}]\xi_{t_1,t_2}\xi_{t_3,t_4} \\ &+ 2(k-1)^2\sum_{t_1,t_2,t_3,t_4}^{p} \mathbb{E}[\tilde{g}_{c,t_1}\tilde{g}_{c,t_3}]\mathbb{E}[\tilde{g}_{c,t_2}\tilde{g}_{c,t_4}]\xi_{t_1,t_2}\xi_{t_3,t_4}\end{aligned}$$

or on a :

$$\sum_{t_1,t_2,t_3,t_4}^{p} \mathbb{E}[\tilde{g}_{c,t_1}\tilde{g}_{c,t_2}]\mathbb{E}[\tilde{g}_{c,t_3}\tilde{g}_{c,t_4}]\xi_{t_1,t_2}\xi_{t_3,t_4} = p^2 - 2pO(p^2r_n^2) + O(p^4r_n^4).$$

On peut aussi remplaçer $\mathbb{E}[\tilde{g}_{c,t_1}\tilde{g}_{c,t_2}\tilde{g}_{c,t_3}\tilde{g}_{c,t_4}]$ par $\mathbb{E}[\tilde{g}_{t_1}\tilde{g}_{t_2}\tilde{g}_{t_3}\tilde{g}_{t_4}]$ en remarquant que c'est le terme dominant parmi la décomposition en seize termes. On obtient donc

$$\begin{aligned}&\mathbb{E}[(\sum_{j=1}^{n}\mathbb{E}[W_{nj}^2|\mathcal{F}_{n,j-1}])^2] \\ =\ & \sum_{j=1}^{n}\mathbb{E}[(\mathbb{E}[W_{nj}^2|\mathcal{F}_{n,j-1}])^2] + 2\sum_{k<j}\mathbb{E}[\mathbb{E}[W_{nj}^2|\mathcal{F}_{n,j-1}]\mathbb{E}[W_{nk}^2|\mathcal{F})_{n,k-1}] \\ =\ & \sum_{j=1}^{n}(j-1)\sum_{t_1,t_2,t_3,t_4}^{p} \mathbb{E}[\tilde{g}_{t_1}\tilde{g}_{t_2}\tilde{g}_{t_3}\tilde{g}_{t_4}]\xi_{t_1,t_2}\xi_{t_3,t_4} + (j-1)^2p^2 + 2(j-1)^2p) \\ +\ & 2\sum_{j-1}^{n}\sum_{k=1}^{j-1}((k-1)\sum_{t_1,t_2,t_3,t_4}^{p} \mathbb{E}[\tilde{g}_{t_1}\tilde{g}_{t_2}\tilde{g}_{t_3}\tilde{g}_{t_4}]\xi_{t_1,t_2}\xi_{t_3,t_4} + (j-1)(k-1)p^2 + 2(k-1)^2p) + o(1) \\ =\ & O(n^3\sum_{t_1,t_2,t_3,t_4}^{p} \mathbb{E}[\tilde{g}_{t_1}\tilde{g}_{t_2}\tilde{g}_{t_3}\tilde{g}_{t_4}]\xi_{t_1,t_2}\xi_{t_3,t_4}) + \frac{n^4}{4}p^2(1+o(1)) + \frac{n^4}{6}p(1+o(1))\end{aligned}$$

On trouve que $\mathbb{E}[(\sum_{j=1}^{n}\mathbb{E}[\widetilde{W}_{nj}^2|\mathcal{F}_{n,j-1}])^2] = 1 + o(1)$, et donc $\mathrm{Var}(\sum_{j=1}^{n}\mathbb{E}[\widetilde{W}_{nj}^2|\mathcal{F}_{n,j-1}]) \longrightarrow 0$

La condition de Lyapunov conditionnelle se traite de la même façon en utilisant les hypothèses D-(b)et D-(f). ■

Preuve du théorème 5.1.

Lemma 7.3 *On a*

$$C'\psi(h) \leq \mathbb{E}K(h^{-1}\|X_1 - x\|) \leq C\psi(h),$$

$$C_1\psi(h) \leq \mathbb{E}K^2(h^{-1}\|X_1 - x\|) \leq C_2\psi(h)$$

et

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} K(h^{-1}\|X_i - x\|)\right) \to 0$$

Proof. La preuve est similaire au lemme 4.4 de Ferraty et Vieu (2006). Les hypothèses sur le noyau K et sur les petites boules de probabilité permettent de dire que le noyau est un noyau de type II et la condition additionnelle est vérifiée ; ce qui permet d'appliquer le lemme 4.4. Concernant la variance, on peut écrire où C et C' sont des constantes :

$$\begin{aligned}
\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} K(h^{-1}\|X_i - x\|)\right) &= \frac{1}{n}\mathrm{Var}\left(K(h^{-1}\|X_k - x\|)\right) \\
&\leq \frac{1}{n}(C\psi(h) - C'\psi^2(h)) \\
&\to 0 \quad .
\end{aligned}$$

On a utilisé le fait que $\psi(h) \to 0$ quand h tend vers 0, ainsi que les deux premières propriétés. ■

Lemma 7.4

$$\sup_{x\in S} |\hat{m}(x,t) - \hat{m}(x^{(k)},t)| = o_{\mathbb{P}}(1).$$

Proof. Pour cela, on utilise le caractère Lipchitzien de K. On peut donc écrire :

$$\begin{aligned}
|K(h^{-1}\|X_i - x\|) - K(h^{-1}\|X_i - x^{(k)}\|)| &\leq \frac{C}{h}\left|\|X_i - x\| - \|X_i - x^{(k)}\|\right| \\
&\leq \frac{C}{h}\|x - x^{(k)}\| \\
&\leq \frac{C}{h}\sup_{x\in S}\|x - x^{(k)}\|.
\end{aligned}$$

Donc,

$$
\begin{aligned}
& |\hat{m}(x,t)-\hat{m}(x^{(k)},t)| \\
= & \frac{1}{n}\left|\sum_{j=1}^{n}\Phi(Y_j,t)\left(\frac{K(h^{-1}\|X_j-x\|)}{\frac{1}{n}\sum_{i=1}^{n}K(h^{-1}\|X_i-x\|)}-\frac{K(h^{-1}\|X_j-x^{(k)}\|)}{\frac{1}{n}\sum_{i=1}^{n}K(h^{-1}\|X_i-x^{(k)}\|)}\right)\right| \\
\leq & \frac{C}{h\psi(h)}\|x-x^{(k)}\|O_{\mathbb{P}}(1).
\end{aligned}
$$

D'où le résultat. ■

Lemma 7.5

$$\sup_{x\in S,t\in\mathbb{R}}|m(x^{(k)},t)-m(x,t)|=o(1)$$

Proof. En effet,

$$
\begin{aligned}
|m(x^{(k)},t)-m(x,t)| & \leq C(t)\|x^{(k)}-x\|_{L^2} \\
& \leq \sup_{t\in\mathbb{R}}C(t)\sup_{x\in S}\|x^{(k)}-x\|_{L^2}.
\end{aligned}
$$

D'où le résultat, en faisant tendre k vers l'infini. ■

Lemma 7.6

$$\sup_{x\in S,t\in\mathbb{R}}|\hat{m}(x^{(k)},t)-m(x^{(k)},t)]|=o_{\mathbb{P}}(1)$$

Proof. On applique le théorème 2 voir aussi le corollaire 2 de Einmahl et Mason (2005). ■

Au vu des différents lemmes, nous obtenons le résultat. ■

Chapitre 3

Projection-based nonparametric goodness-of-fit testing with functional covariates

3.1 Introduction

Consider a sample of independent copies $(U_1, X_1), \cdots, (U_n, X_n)$ of (U, X) where U is a real-valued random variable and X is a square-integrable random function defined on the unit interval. The problem we investigate herein is the test of the hypothesis

$$H_0 : \mathbb{E}(U|X) = 0 \quad \text{almost surely (a.s.)} \tag{3.1.1}$$

against the nonparametric alternative $\mathbb{P}[\mathbb{E}(U|X) = 0] < 1$. We consider two cases : (a) U is directly observed ; and (b) U is not observed and is estimated as a residual of a parametric model for functional covariates and scalar responses.

There has been substantial recent work on the theoretical study of the functional data analysis. The monographs of Ramsay and Silverman (2002, 2005) and Ferraty (2011) provide a comprehensive landscape of the importance of the statistical methods for functional data. Estimation and prediction with functional covariates received substantial attention in the literature : for example by Ferraty and Vieu (2006), Cai and Hall (2006), Hall and Horowitz (2007), Crambes, Kneip and Sarda (2008), Yao and Müller (2010) and the references therein.

The goodness-of-fit problem we address seems to be much less explored. There is a large literature on model checks like (3.1.1) against nonparametric alternatives when X takes values in a finite-dimension space, see for instance Härdle and Mammen (1993), Stute (1997), Horowitz and Spokoiny (2001), Guerre and

Lavergne (2005). In the case of functional covariate X, much little work was accomplished for testing against general types of alternatives. To our best knowledge, the only contribution considering the problem of testing H_0 against nonparametric alternatives in the cases (a) and (b) is the recent paper of Delsol, Ferraty and Vieu (2011) who extend the idea of Härdle and Mammen (1993) to the functional covariate case. However, their results are derived under some strong assumptions, like for instance the assumptions on the rates of convergence of the so-called small ball probabilities and the law of the covariate X that are supposed to be known. It is not clear how the test of Delsol, Ferraty and Vieu (2011) could be easily applied in practice, for instance for testing the goodness-of-fit of the functional linear model. Some more substantial work was done for testing for no effect in a functional linear model, see Cardot, Ferraty, Mas and Sarda (2003), Cardot, Goia and Sarda (2007), or for testing the functional linear model against quadratic alternatives, see Horvàth and Reeder (2011). By construction, such procedures are not able to detect general departures from the null hypothesis.

The test we introduce herein is based on a dimension reduction idea used by Lavergne and Patilea (2008) in a finite dimension setup. Our test is able to detect *nonparametric* alternatives, including the polynomial ones. The variable U could be heteroscedastic and we do not require the conditional variance of U given X to be known. We do no require the law of the covariate X to be given or to be of a certain type, like for instance Gaussian. The test could be implemented quite easily and performs well in simulations and real data applications.

The paper is organized as follows. In section 3.2 we introduce the main notation and we derive a fundamental lemma for our approach. This lemma shows that checking condition (3.1.1) is equivalent to checking the nullity of the conditional expectation of U given a sufficiently rich set of projections of X on elements of norm 1 from finite-dimension subspaces of $L^2[0,1]$. Next, the idea is to search in finite-dimension subspaces of $L^2[0,1]$ a least favorable element of norm 1 and to check the nullity of the conditional expectation of U given the scalar product between X and the selected least favorable direction. In section 3.3 we introduce the test statistic for testing of no-effect of X on U when U is observed. Our statistic is a quadratic form, based on *univariate* kernel smoothing, that behaves asymptotically like a *standard normal* random variable under H_0. We prove that, under mild integrability or boundedness assumptions, the induced test is consistent against *any* type of fixed alternatives and against sequences of directional alternatives approaching the null hypothesis at a suitable rate. The allowed rates are almost the same as those obtained in parametric model checks based on kernel smoothing with *univariate* covariate, see for instance Guerre and Lavergne (2005) or Lavergne and Patilea (2008). In section 3.4 we apply our projection-based approach for nonparametric checks of the functional regression models and we will focus on the linear functional model. The methodology we propose can also be applied for goodness-of-fit testing of other functional regression models, like for instance the partial functional linear model, see Ramsay and Silverman (2005) chapter 10, and the generalized functional linear models introduced by Müller and Stadtmüller (2005). In the functional regression case the variable U is the unobserved error term of the regression model and hence the test statistic is based on the estimated residuals. We still obtain standard normal critical values and consistency against nonparametric alternatives, fixed or approaching the null hypothesis. However, more restrictive conditions on the bandwidths are required due to the estimation of the slope of the functional linear model. This induces restrictions on the rate the directional alternatives may approach the null hypothesis. More difficult the estimation of the slope parameter is, slower the rate the directional alternative approach the null hypothesis should be. For estimating the slope parameter in the functional linear regression model we will focus on the standard approach based on functional principal component analysis. In section 3.5 an empirical study is reported. First, a wild bootstrap procedure is proposed as a means to approximate the critical values of the test statistic with finite samples. Then, the results of a simulation study are briefly explained. The conclusion is that the test works well in practice. Under the null, the level is quite well respected and the power is more than acceptable even in the comparison with parametric tests. The proposed test is consistent under general alternatives. Some advices and comments are provided about the choice of the parameters involved in the new test. The test is applied to test the goodness-of-fit of the functional linear model and the functional quadratic model for the Tecator data set. Both models are rejected which indicates that more flexible models should be considered, like for instance the semiparametric index models introduced by Chen, Hall and Müller (2011). The proofs of our theoretical results are relegated to the appendix.

3.2 Dimension reduction in nonparametric testing

Let us introduce some notation. For any $p \geq 1$, let $\mathcal{S}^p = \{\gamma \in \mathbb{R}^p : \|\gamma\| = 1\}$ denote the unit hypersphere in $\mathbb{R}^p$. Let $L^2[0,1]$ be the space of the square-integrable real-valued functions defined on the unit interval $\langle \cdot, \cdot \rangle$ denote the inner product in $L^2[0,1]$, that is for any $X_1, X_2 \in L^2[0,1]$

$$\langle X_1, X_2 \rangle = \int_0^1 X_1(t) X_2(t) dt.$$

Let $\|\cdot\|_{L^2}$ be the associated norm. Hereafter $\mathcal{R} = \{\rho_1, \rho_2, \cdots\}$ will be an arbitrarily fixed orthonormal basis of the function space $L^2[0,1]$, that is $\langle \rho_i, \rho_j \rangle = \delta_{ij}$. Then the predictor process X can be expanded into

$$X(t) = \sum_{j=1}^{\infty} x_j \rho_j(t), \tag{3.2.2}$$

where the random coefficients x_j are given by $x_j = \langle X, \rho_j \rangle$. For a fixed positive integer p, $X^{(p)} \in L^2[0,1]$ will be the projection of X on the subspace generated by the first p elements of the basis $\mathcal{R}$, that is

$$X^{(p)}(t) = \sum_{j=1}^{p} x_j \rho_j(t).$$

Let us notice that $\|X^{(p)}\|_{L^2}$ coincides with the Euclidean norm of the vector $(x_1, \cdots, x_p)$ in $\mathbb{R}^p$. By abuse we also identify $X^{(p)}$ with the $p-$dimension random vector $(x_1, \cdots, x_p)$. On the other hand, for any integer $p > 1$ and non random vector $\gamma = (\gamma_1, \cdots, \gamma_p) \in \mathbb{R}^p$, we consider by abuse γ an element in $L^2[0,1]$ with $(\gamma_1, \cdots, \gamma_p, 0, 0, \cdots)$ the coefficients of its expansion and hence $\langle X, \gamma \rangle = \langle X^{(p)}, \gamma \rangle = \sum_{i=1}^{p} x_j \gamma_j$. In the following we will also use $\beta = \sum_{j=1}^{\infty} b_j \rho_j(t)$ to denote a non random element of $L^2[0,1]$.

Our approach relies on the following lemma, an extension of Lemma 2.1 of Lavergne and Patilea (2008) and Theorem 1 in Bierens (1990) to Hilbert space-valued conditioning random variables. The result shows that for checking nullity of a conditional expectation, it is equivalent to consider expectations conditional on X and expectations conditional on $L^2[0,1]$ projections of X on a sufficiently rich set of directions.

Lemma 2.1 *Let $X \in L^2[0,1]$ and $Z \in \mathbb{R}$ be random variables. Assume that $\mathbb{E}|Z| < \infty$ and $\mathbb{E}(Z) = 0$.*

(A) The following statements are equivalent :

1. $\mathbb{E}(Z \mid X) = 0$ *a.s.*
2. $\mathbb{E}(Z \mid \langle X, \beta \rangle) = 0$ *a.s.* $\forall \beta \in L^2[0,1]$ *with* $\|\beta\|_{L^2} = 1$.
3. *for any integer* $p \geq 1$, $\mathbb{E}(Z \mid \langle X, \gamma \rangle) = 0$ *a.s.* $\forall \gamma \in \mathcal{S}^p$.
4. *for any integer* $p \geq 1$, $\mathbb{E}(Z \mid X^{(p)}) = 0$ *a.s.*

(B) Suppose in addition that for any positive real number s,

$$\mathbb{E}(|Z| \exp\{s\|X\|\}) < \infty. \tag{3.2.3}$$

If $\mathbb{P}[\mathbb{E}(Z \mid X) = 0] < 1$, then there exists a positive integer $p_0 \geq 1$ such that for any integer $p > p_0$, the set

$$\{\gamma \in \mathcal{S}^p : \mathbb{E}(Z \mid \langle X, \gamma \rangle) = 0 \ a.s.\}$$

has Lebesgue measure zero on the unit hypersphere $\mathcal{S}^p$ and is not dense.

Point (A) is a cornerstone for proving the behavior of our test under the null and the alternative hypothesis. Point (B) shows that in applications it will not be difficult to find directions γ able to reveal the failure of the null hypothesis (3.1.1). Under the additional assumption (3.2.3) such directions represent almost all the points on the unit hyperspheres $\mathcal{S}^p$, provided p is sufficiently large. The assumption (3.2.3) is not restrictive for testing purposes. Indeed, if X does not satisfy condition (3.2.3), it suffices to transform X into some variable $W \in L^2[0,1]$ such that the $\sigma-$field generated by W is the same as the one generated by X and the variable W satisfies condition (3.2.3).* Clearly, when U is the error term in some functional regression model for which one wants to check the goodness-of-fit, one should use a transformation of X only *after* estimating the errors in the model.

The following new formulations of H_0 are direct consequences of Lemma 2.1-(A).

Corollary 2.2 *Consider a real-valued random variable U such that $\mathbb{E}|U| < \infty$. Let $\omega(\beta, t)$, $\beta \in L^2[0,1]$ and $t \in \mathbb{R}$, be a real-valued function such that $\omega(\beta, \langle X, \beta\rangle) > 0$ for all $\|\beta\|_{L^2} = 1$. For any $p \geq 1$, let $w_p(\gamma, t)$, $\gamma \in \mathbb{R}^p$ and $t \in \mathbb{R}$, be a real-valued function such that $w_p(\gamma, \langle X, \gamma\rangle) > 0$ for all $\|\gamma\| = 1$. The following statements are equivalent :*

1. *The null hypothesis (3.1.1) holds true.*
2. $$\max_{\beta \in L^2[0,1],\ \|\beta\|_{L^2}=1} \mathbb{E}\left[U \mathbb{E}\left(U | \langle X, \beta\rangle\right) \omega(\beta, \langle X, \beta\rangle)\right] = 0. \tag{3.2.4}$$
3. *for any $p \geq 1$ and any set $B_p \subset \mathcal{S}^p$ with strictly positive Lebesgue measure on the unit hypersphere $\mathcal{S}^p$,*
$$\max_{\gamma \in B_p} \mathbb{E}\left[U \mathbb{E}\left(U | \langle X, \gamma\rangle\right) w_p(\gamma, \langle X, \gamma\rangle)\right] = 0. \tag{3.2.5}$$

3.3 Testing the effect of a functional covariate

We introduce a general approach for nonparametric testing of the effect of a functional covariate X on a real-valued random variable U. For simplicity, here we assume that $\mathbb{E}(U) = 0$, the nonzero mean case is contained in the setup considered in section 3.4 below. Our approach is based on Corollary 2.2-(3) and *univariate* kernel smoothing. In this way we avoid the problem of smoothing in infinite-dimension, in particular we avoid using the small ball function required in the kernel regression

*. For instance, given $X = \sum_{j\geq 1} x_j \rho_j$, one may build $w_j = a_j \arctan(x_j)$, where a_j are non random such that $\sum_{j\geq 1} a_j^2 < \infty$ and may use the bounded random function $W = \sum_{j\geq 1} w_j \rho_j \in L^2[0,1]$ (bounded means $\|W\|$ is a bounded random variable) instead of X in the conditioning.

with functional covariates, see Ferraty and Vieu (2006), Delsol, Ferraty and Vieu (2011).

To avoid handling denominators close to zero, we set the weight function $\omega(\gamma, \cdot)$ in Corollary 2.2 equal to the density of $\langle X, \gamma\rangle$, denoted by $f_\gamma(\cdot)$, which is assumed to exist for any γ. For any $\gamma \in \mathbb{R}^p$, let

$$Q(\gamma) = \mathbb{E}\{U\, \mathbb{E}[U \mid \langle X, \gamma\rangle] f_\gamma(\langle X, \gamma\rangle)\} = \mathbb{E}\{\mathbb{E}^2[U \mid \langle X, \gamma\rangle] f_\gamma(\langle X, \gamma\rangle)\}.$$

For any $p \geq 1$, let $B_p \subset \mathcal{S}^p$ be a set with strictly positive Lebesgue measure in $\mathcal{S}^p$. By Corollary 2.2, the null hypothesis (3.1.1) holds true if and only if

$$\forall p \geq 1, \quad \max_{\gamma \in B_p} Q(\gamma) = 0. \tag{3.3.1}$$

3.3.1 The test statistic

In view of equation (3.3.1), our goal is to estimate $Q(\gamma)$. With at hand a sample of (U, X), define

$$Q_n(\gamma) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} U_i U_j \frac{1}{h} K_h(\langle X_i - X_j, \gamma\rangle), \quad \gamma \in \mathcal{S}^p,$$

where $K_h(\cdot) = K(\cdot/h)$, where $K(\cdot)$ is a kernel and h a bandwidth. In the case of finite dimension covariates, the function $\gamma \mapsto Q_n(\gamma)$ is the statistic considered by Lavergne and Patilea (2008), see also Bierens (1990). For fixed p and $\gamma \in \mathcal{S}^p$, it is well-known that $Q_n(\gamma)$ has asymptotic centered normal distribution with rate $nh^{1/2}$ under H_0; see for instance Guerre and Lavergne (2005). We will show that the asymptotic normal distribution is preserved even when p grows at a suitable rate with the sample size. On the other hand, Lemma 2.1-(B) indicates that if p is large enough, the maximum of $Q(\gamma)$ over γ stays away from zero when H_0 fails.

For a fixed p the statistic $Q_n(\gamma)$ is expected to be close to $Q(\gamma)$ uniformly in γ. Then a natural idea would be to build a test statistic using the maximum of $Q_n(\gamma)$ with respect to γ. However, there is an additional difficulty one faces in the functional data framework since then one has to let p grow to infinity with the sample size, and hence the closeness between $Q_n(\gamma)$ and $Q(\gamma)$ requires a more careful investigation. On the other hand, like in the finite dimension covariate case, under H_0 one expects $Q_n(\gamma)$ to converge to zero for any p and γ and thus the objective function of the maximization problem to be flat.

We will choose a direction γ as the least favorable direction for the null hypothesis H_0 obtained from a penalized criterion based on a standardized version of $Q_n(\gamma)$. Lavergne and Patilea (2008) and Bierens (1990) considered this idea using $Q_n(\gamma)$.

Here we use a standardized version of $Q_n(\gamma)$. More precisely, fix some $\beta_0 \in L^2[0,1]$ that could be interpreted as an initial *guess* of an unfavorable direction for H_0. Let b_{0j}, $j \geq 1$, be the coefficients in the expansion of β_0 in the basis $\mathcal{R}$. For any given $p \geq 1$ such that $\sum_{j=1}^{p} b_{0j}^2 > 0$, let

$$\gamma_0^{(p)} = (b_{01}, \cdots, b_{0p}) / \sqrt{\sum_{j=1}^{p} b_{0j}^2} \; .$$

Let $\widehat{v}_n^2(\cdot)$ be an estimate of the variance of $nh^{1/2}Q_n(\cdot)$. Given $B_p \subset \mathcal{S}^p$ with strictly positive Lebesgue measure in $\mathcal{S}^p$ that contains $\gamma_0^{(p)}$, the least favorable direction γ for H_0 is defined as

$$\widehat{\gamma}_n = \arg\max_{\gamma \in B_p} \left[nh^{1/2} Q_n(\gamma)/\widehat{v}_n(\gamma) - \alpha_n \mathbb{I}_{\left\{\gamma \neq \gamma_0^{(p)}\right\}} \right] , \tag{3.3.2}$$

where $\mathbb{I}_A$ is the indicator function of a set A, and α_n, $n \geq 1$ is a sequence of positive real numbers increasing to infinity at an appropriate rate that depends on the sample size and the rates of h and p and that will be made explicit below. Let us notice that the maximization used to define $\widehat{\gamma}_n \in \mathcal{S}^p$ is a finite dimension optimization problem. The choice of β_0, and thus of $\gamma_0^{(p)}$, is theoretically irrelevant, it does not affect the asymptotic critical values and the consistency results. However, in practice the choice of β_0 could be related to a priori information of the practitioner on a class of alternatives, like for instance the class of functions depending only on $\langle X, \beta_0 \rangle$. The empirical investigation we report in section 3.5 suggests that working with a standardized version of $Q_n(\gamma)$ simplifies the choice of α_n in applications.

We will prove that with suitable rates of increase for α_n and p and decrease for h, the probability of the event $\{\widehat{\gamma}_n = \gamma_0^{(p)}\}$ tends to 1 under H_0. Hence $Q_n(\widehat{\gamma}_n)/\widehat{v}_n(\widehat{\gamma})$ behaves asymptotically like $Q_n(\gamma_0^{(p)})/\widehat{v}_n(\gamma_0^{(p)})$, even when p grows with the sample size. Therefore the test statistic we consider is

$$T_n = nh^{1/2} \frac{Q_n(\widehat{\gamma}_n)}{\widehat{v}_n(\widehat{\gamma}_n)} \; . \tag{3.3.3}$$

We will show that an asymptotic α-level test is given by $\mathbb{I}\,(T_n \geq z_{1-\alpha})$, where $z_{1-\alpha}$ is the $(1-\alpha)$-th quantile of the standard normal distribution.

3.3.2 Estimating the variance

To find the direction $\widehat{\gamma}_n$ and to build the test statistics (3.3.3), we need to estimate in some way the variance of $nh^{1/2}Q_n(\gamma)$. The approach that is expected not to inflate the variance estimate under the alternatives and thus to guarantee better power would involve the estimations of the conditional variance of $nh^{1/2}Q_n(\gamma)$ given X_i's which writes

$$\tau_n^2(\gamma) = \frac{2}{n(n-1)h}\sum_{j\neq i}\sigma_p^2(X_i^{(p)})\sigma_p^2(X_j^{(p)})K_h^2(\langle X_i - X_j, \gamma\rangle) \ , \tag{3.3.4}$$

where $\sigma_p^2(X^{(p)}) = Var[U \mid X^{(p)}]$. An estimator can be easily obtained by replacing $\sigma_p^2(\cdot)$ with an estimate in the last expression. In theory, a good solution would be to use a nonparametric estimate of the $p-$variate function $\sigma_p^2(\cdot)$, but this is practically infeasible given that it expected to let p grow with the sample size. A simple and convenient solution with high-dimension covariates is then

$$\widehat{\tau}_n^2(\gamma) = \frac{2}{n(n-1)h}\sum_{j\neq i}U_i^2U_j^2K_h^2(\langle X_i - X_j, \gamma\rangle). \tag{3.3.5}$$

Since, under the null hypothesis $\widehat{\gamma}_n = \gamma_0^{(p)}$ with probability tending to 1, a first variance estimator we propose is

$$\widehat{v}_n^2(\widehat{\gamma}_n) = \widehat{v}_n^2(\widehat{\gamma}_n, \gamma_0^{(p)}) = \min\left(\widehat{\tau}_n^2(\widehat{\gamma}_n), \widehat{\tau}_n^2(\gamma_0^{(p)})\right). \tag{3.3.6}$$

On the other hand, let us notice that $\tau_n^2(\gamma_0^{(p)}) - \mathbb{E}[\tau_n^2(\gamma_0^{(p)})]$ is expected to converge to zero. Moreover, under the null hypothesis

$$\mathbb{E}[\tau_n^2(\gamma_0^{(p)})] = \mathbb{E}[\mathbb{E}[\tau_n^2(\gamma_0^{(p)}) \mid \langle X_1, \gamma_0^{(p)}\rangle, \cdots, \langle X_n, \gamma_0^{(p)}\rangle]]$$
$$= \mathbb{E}\left\{\frac{2}{n(n-1)h}\sum_{j\neq i}Var[U_i \mid \langle X_i, \gamma_0^{(p)}\rangle]\ Var[U_j \mid \langle X_j, \gamma_0^{(p)}\rangle]K_h^2\left(\langle X_i - X_j, \gamma_0^{(p)}\rangle\right)\right\},$$

and $0 < \underline{\sigma}^2 \leq Var[U \mid \langle X, \gamma_0^{(p)}\rangle] \leq \overline{\sigma}^2 < \infty$. Next, notice that the conditional variance of U given $\langle X, \beta_0\rangle$ is the same under H_0 and under any alternative that depends only on $\langle X, \beta_0\rangle$. Finally, notice that in any case $\mathbb{E}(U^2) \geq \mathbb{E}[Var(U \mid \langle X, \beta_0\rangle)]$. All these facts suggest that a compromise for estimating the variance of $nh^{1/2}Q_n(\beta_0^{(p)})$ would be

$$\widehat{v}_n^2 = \widehat{v}_n^2(\gamma_0^{(p)}) = \frac{2}{n(n-1)h}\sum_{j\neq i}\widehat{\sigma}^2_{\gamma_0^{(p)}}(\langle X_i, \gamma_0^{(p)}\rangle)\ \widehat{\sigma}^2_{\gamma_0^{(p)}}(\langle X_j, \gamma_0^{(p)}\rangle)K_h^2\left(\langle X_i - X_j, \gamma_0^{(p)}\rangle\right), \tag{3.3.7}$$

where $\widehat{\sigma}^2_{\gamma_0^{(p)}}(\cdot)$ is some nonparametric estimate of the univariate function $\sigma^2_{\gamma_0^{(p)}}(t) = Var(U \mid \langle X, \gamma_0^{(p)}\rangle = t)$ satisfying the condition

$$\sup_{1\leq i\leq n}\left|\frac{\widehat{\sigma}^2_{\gamma_0^{(p)}}(\langle X_i, \gamma_0^{(p)}\rangle)}{\sigma^2_{\gamma_0^{(p)}}(\langle X_i, \gamma_0^{(p)}\rangle)} - 1\right| = o_{\mathbb{P}}(1)\ , \tag{3.3.8}$$

Different nonparametric estimators can be used, for instance a kernel estimator like in Lavergne and Patilea (2008). We will prove below that both variance estimators (3.3.6) and (3.3.7) guarantee the standard normal asymptotic critical values and consistency of our test. In simulations, better power under the alternative was obtained when using the variance estimator (3.3.7). The drawback of this estimator is the computational cost and the choice of an additional bandwidth for the estimate $\widehat{\sigma}^2_{\gamma_0^{(p)}}(\cdot)$. However, common choices of this bandwidth work well in practice.

3.3.3 Behavior under the null hypothesis

Let us introduce a first set of assumptions. Below $0_p \in \mathbb{R}^p$ denotes the null vector of dimension p. Moreover, $\mathcal{F}[\cdot]$ denotes the Fourier transform, cf. Rudin (1987).

Assumption D

(a) The random vectors $(U_1, X_1), \ldots, (U_n, X_n)$ are independent draws from the random vector $(U, X) \in \mathbb{R} \times L^2[0,1]$ that satisfies $\mathbb{E}|U|^m < \infty$ for some $m > 11$.

(b) $\exists\ \underline{\sigma}^2$ and $\overline{\sigma}^2$ such that $0 < \underline{\sigma}^2 \leq Var(U \mid X) \leq \overline{\sigma}^2 < \infty$ almost surely.

(c) The sets $B_p \subset \mathcal{S}^p$, $p \geq 1$ appearing in (3.3.2) are such that :

(i) there exist constants $C_1, \delta > 0$ (independent of n and p) such that $\forall p \geq 1$ and $\forall \gamma \in B_p$, the variable $\langle X, \gamma \rangle$ admits a density $f_\gamma(\cdot)$ and

$$C_1^{-1} \leq \int_{\mathbb{R}} \{f_\gamma^2 + f_\gamma^{2+\delta}\mathbb{I}(f_\gamma > 1)\} \leq C_1;$$

(ii) there exists $C_2, \epsilon > 0$ such that $\int_{|x|\leq\epsilon} |\mathcal{F}[f_\gamma]|^2(x)dx \geq C_2$, $\forall p \geq 1$, $\forall \gamma \in B_p$;

(iii) the initial 'guess' β_0 satisfies the condition : $\exists C_3$ such that $f_{\gamma_0^{(p)}} \leq C_3$, $\forall p \geq 1$.

(iv) $B_p \times 0_{p'-p} \subset B_{p'}$, $\forall 1 \leq p < p'$.

Assumption K

(a) The kernel K is a symmetric continuous density of bounded variation with strictly positive Fourier transform on the real line.

(b) $h \to 0$ and $(nh^2)^\alpha / \ln n \to \infty$ for some $\alpha \in (0,1)$.

(c) $p \geq 1$ increases to infinity with n and there exists a constant $\lambda > 0$ such that $p \ln^{-\lambda} n$ is bounded.

Let us comment on these assumptions. The bounded variation of K, in particular this means K is bounded, is a very mild condition that allows to easily bound covering numbers of families of functions indexed by γ. Continuity and bounded variation guarantee that K can be recovered by inverse Fourier transform. The role of technical assumption of positive Fourier, that is satisfied by triangular, normal, logistic, Student, or Laplace densities, will be explained below. In Assumption K-(c), it is also possible to let p to grow with the sample size at a polynomial rate, instead of the logarithmic rate. However, we will see below that, in theory, this could induce a loss of power for our test. There is a trade off between the moment conditions one imposes for U and the range of rates allowed for the bandwidth and the growth rate for p : higher moments will be needed for wider ranges and faster rates for p. For bandwidths and p satisfying Assumption K-(b,c) it suffices to take $m > 11$ in Assumption D-(a) ; see the proof of Lemma 3.1. Let us notice that Assumption D-(b) implies that $\forall p \geq 1$, $0 < \underline{\sigma}^2 \leq \mathbb{E}(U^2 \mid X^{(p)}) \leq \overline{\sigma}^2 < \infty$ almost surely. Finally, let us comment on Assumption D-(c). On one hand, a key issue in the proof of Lemma 3.1 below and some of the subsequent proofs will be to control the rate of $\mathbb{E}[h^{-1}K_h(\langle X_1 - X_2, \gamma\rangle)]$ uniformly in $\gamma \in B_p$ as p grows and h decreases with the sample size. To reduce technicalities, we choose the convenient solution that consists in trying to bound this quantity by a constant. Using the Fourier transform and Plancherel theorem, this is guaranteed by a condition like $\int_{\mathbb{R}} f_\gamma^2 \leq C_1, \forall \gamma \in B_p$. In the proofs for the functional linear model we have to strengthen this condition and add $\int_{\mathbb{R}} f_\gamma^{2+\delta}\mathbb{I}(f > 1) \leq C_1$, $\forall \gamma \in B_p$, for some arbitrary small $\delta > 0$. Such sufficient conditions could be easily achieved for instance if the coefficients x_j of the expansion of X are independent. Then it suffices to fix some $k \geq 1$ such that the density of x_k is bounded and some small c independent of p and to take $B_p = \{(\gamma_1, \cdots, \gamma_k, \cdots, \gamma_p) \in \mathcal{S}^p : |\gamma_k| \geq c\}$. This simple idea could be useful in many other cases than the one of independent coefficients x_j. On the other hand, we have to keep the variance estimate in the denominator of the test statistic (3.3.3) away from zero. For this we have to ensure that $\mathbb{E}[h^{-1}K_h^2(\langle X_1 - X_2, \gamma\rangle)]$ is bounded away from zero uniformly in $\gamma \in B_p$ as p grows and h decreases with the sample size. One easy way to ensure this is to use again the Fourier transform properties, the positiveness of $\mathcal{F}[K]$ and to impose the positive uniform lower bound for the integral of square of $\mathcal{F}[f_\gamma]$ in a neighborhood of the origin, which necessarily induces a uniform lower bound for $\int_{\mathbb{R}} f_\gamma^2$. Assumptions D-(c)(iii) will complete the sufficient conditions for deriving standard normal critical values using the central limit theorem for $U-$statistics of Guerre and Lavergne (2005, Lemma 2). To summarize, the choice of β_0 and B_p will be decided in the applications and will also depend on the law of X and the choice of the basis $\mathcal{R}$. In view of our extensive simulation experiment, we argue that the choice of B_p is not an issue in applications, one can confidently perform the optimization on the whole hypersphere $\mathcal{S}^p$. Finally, the condition $B_p \times 0_{p'-p} \subset B_{p'}$, $\forall p < p'$, is a mild technical condition that combined with Lemma 2.1-(A) greatly simplifies the proof of the consistency of our test.

The first step is the study of the behavior of the process $Q_n(\gamma)$, $\gamma \in B_p$, under H_0 when p is allowed to increase with the sample size.

Lemma 3.1 *Under Assumptions D and K and if H_0 holds true,*

$$\sup_{\gamma \in B_p \subset \mathcal{S}^p} |Q_n(\gamma)| = O_{\mathbb{P}}(n^{-1} h^{-1/2} p^{3/2} \ln n).$$

Moreover, if $\widehat{\tau}_n^2(\gamma)$ is the estimate defined in equation (3.3.5),

$$\sup_{\gamma \in B_p \subset \mathcal{S}^p} \{1/\widehat{\tau}_n^2(\gamma)\} = O_{\mathbb{P}}(1).$$

If in addition condition (3.3.8) holds true, $1/\widehat{v}_n^2 = O_{\mathbb{P}}(1)$ with $\widehat{v}_n^2$ defined in (3.3.7).

We now describe the behavior of $\widehat{\gamma}_n$ under H_0. A suitable rate α_n will make $\widehat{\gamma}_n$ to be equal to $\gamma_0^{(p)}$ with high probability. Under the null, α_n has to grow to infinity sufficiently fast to render the probability of the event $\{\widehat{\gamma}_n = \gamma_0^{(p)}\}$ close to 1. We will see below that, for better detection of alternative hypothesis, α_n should grow as slow as possible. Indeed, slower rates for α_n will allow the selection of directions $\hat{\gamma}_n$ that could be better suited than $\gamma_0^{(p)}$ for revealing the departure from the null hypothesis. The rate of p is also involved in the search of a trade-off for the rate of α_n : larger p renders slower the rate of uniform convergence to zero of $Q_n(\gamma)$, $\gamma \in B_p$, and hence requires larger α_n.

Lemma 3.2 *Under Assumptions D, K, and condition (3.3.8) if the variance estimator is the one defined in (3.3.7), for a positive sequence α_n, $n \geq 1$ such that $\alpha_n / \{p^{3/2} \ln n\} \to \infty$,*

$$\mathbb{P}(\widehat{\gamma}_n = \gamma_0^{(p)}) \to 1, \qquad \textit{under } H_0.$$

The following result shows that the asymptotic critical values of our test statistic are standard normal.

Theorem 3.3 *Under the conditions of Lemma 3.2 and if the hypothesis H_0 in (3.1.1) holds true, the test statistic T_n converges in law to a standard normal. Consequently, the test given by $\mathbb{I}(T_n \geq z_{1-a})$, with z_a the $(1-a)-$quantile of the standard normal distribution, has asymptotic level a.*

3.3.4 The behavior under the alternatives

First let us give an intuition on the reason why our test is consistent. Consider the alternative hypothesis
$$H_1: \mathbb{P}[\mathbb{E}(U \mid X) = 0] < 1.$$
The way the statistic T_n is constructed guarantees the consistency of our test against H_1. Indeed, we can write

$$\begin{aligned} T_n &= \frac{nh^{1/2}Q_n(\widehat{\gamma}_n)}{\widehat{v}_n(\widehat{\gamma}_n)} \\ &= \max_{\gamma \in B_p} \left\{ nh^{1/2}Q_n(\gamma)/\widehat{v}_n(\gamma) - \alpha_n \mathbb{I}_{\{\gamma \neq \gamma_0^{(p)}\}} \right\} + \alpha_n \mathbb{I}_{\{\widehat{\gamma}_n \neq \gamma_0^{(p)}\}} \\ &\geq \max_{\gamma \in B_p} \frac{nh^{1/2}Q_n(\gamma)}{\widehat{v}_n(\gamma)} - \alpha_n \geq \frac{nh^{1/2}Q_n(\widetilde{\gamma})}{\widehat{v}_n(\widetilde{\gamma})} - \alpha_n, \quad \forall \widetilde{\gamma} \in B_p \subset \mathcal{S}^p, \end{aligned} \tag{3.3.9}$$

with $\widehat{v}_n(\gamma)$ equal to $\widehat{\tau}_n(\gamma)$ defined in (3.3.5) or $\widehat{v}_n(\gamma)$ defined like in (3.3.7). Since $\mathbb{E}(U^2 \mid X) \geq \underline{\sigma}^2$ and $Var(U \mid \langle X, \gamma \rangle) \geq \underline{\sigma}^2$, $\forall \gamma$, it is clear that $\sup_\gamma \{1/\widehat{v}_n(\gamma)\} = O_{\mathbb{P}}(1)$ for both variance estimates introduced above. On the other hand, from Lemma 2.1, there exists p_0 and $\widetilde{\gamma} \in B_{p_0}$ such that the expectation of $Q_n(\widetilde{\gamma})$ stays away from zero as the sample size grows to infinity and h decrease to zero. Finally, for any $p > p_0$ and any n and h, clearly $\max_{\gamma \in B_p} Q_n(\gamma) \geq Q_n(\widetilde{\gamma})$, because $B_{p_0} \times 0_{p-p_0} \subset B_p$. All these facts show why our test is omnibus, that is consistent against nonparametric alternatives, provided that $p \to \infty$.

To formalize the consistency result, let us fix some real-valued function $\delta(X)$ such that $\mathbb{E}[\delta(X)] = 0$ and $0 < \mathbb{E}[\delta^4(X)] < \infty$, and some sequence of real numbers r_n that could decrease to zero (the case $r_n \equiv 1$ is also included). Consider the sequence of alternatives

$$H_{1n}: \ U = U^0 + r_n \delta(X), \quad n \geq 1, \quad \text{with } \mathbb{E}(U^0 \mid X) = 0. \tag{3.3.10}$$

We show below that such directional alternatives can be detected as soon as $r_n^2 nh^{1/2}/\alpha_n$ tends to infinity. This is exactly the same condition as in Lavergne and Patilea (2008). However, in the functional data framework, to obtain the convenient standard normal critical values, we need $1/\alpha_n = o(p^{-3/2} \ln^{-1} n)$. Hence, the rate r_n at which the alternatives H_{1n} tend to the null hypothesis should satisfy $r_n^2 nh^{1/2}/\{p^{3/2} \ln n\} \to \infty$.

Theorem 3.4 *Suppose that*

(a) Assumption D holds true with U replaced by U^0;

(b) Assumption K is satisfied, and so is the condition (3.3.8) if the variance estimator is the one defined in (3.3.7);

(c) $\alpha_n/\{p^{3/2}\ln n\} \to \infty$ *and* r_n, $n \geq 1$ *is such that* $r_n^2 n h^{1/2}/\alpha_n \to \infty$ *;*

(d) $\mathbb{E}[\delta(X)] = 0$ *and* $0 < \mathbb{E}[\delta^4(X)] < \infty$*.*

Then the test based on T_n *is consistent against the sequence of alternatives* H_{1n} *if there exists* $p \geq 1$ *and* $\widetilde{\gamma} \in B_p$ *such that* $\mathbb{P}(\mathbb{E}[\delta(X) \mid \langle X, \widetilde{\gamma}\rangle] = 0) < 1$ *and one of the following conditions is satisfied :*

1. *the function* $\mathbb{E}[\delta(X) \mid \langle X, \widetilde{\gamma}\rangle = \cdot] f_{\widetilde{\gamma}}(\cdot)$ *is bounded;*
2. *the Fourier transform of* $\mathbb{E}[\delta(X) \mid \langle X, \widetilde{\gamma}\rangle = \cdot] f_{\widetilde{\gamma}}(\cdot)$ *is integrable on* $\mathbb{R}$*.*

Conditions (1) or (2) in Theorem 3.4 impose mild restrictions on the function $\delta(\cdot)$ and allow for rather simple proofs, but clearly many alternative sets of technical conditions could be considered. Let us recall that the existence of p and $\widetilde{\gamma} \in B_p$ such that $\mathbb{P}[\delta(X) \mid \langle X, \widetilde{\gamma}\rangle] = 0) < 1$ is guaranteed by Lemma 2.1.

3.4 Testing the goodness-of-fit of parametric models

Here we apply our projection-based testing methodology for testing the goodness-of-fit of the functional linear regression model against *nonparametric* alternatives satisfying mild technical conditions. Hence we provide a simple goodness-of-fit procedure for a widely used model. To the best of our knowledge, our results are completely new in the functional regression framework.

Let U be a real-valued random variable and X be a random variable with values in $L^2[0,1]$. The model we want to test is the functional linear model defined by

$$Y = a + \langle b, X\rangle + U,$$

where $b \in L^2[0,1]$ and $a \in \mathbb{R}$ are unknown parameters. The null hypothesis is

$$H_0 : \ \mathbb{E}(U|X) = 0 \quad \text{a.s.} \tag{3.4.11}$$

Like in Assumption D we consider that $(U_1, X_1), \cdots, (U_n, X_n)$ are independent copies of (U, X), but now the observations are $(Y_1, X_1), \cdots, (Y_n, X_n)$. Hence the error term U has to be estimated in a preliminary step from the estimates of the parameters a and b. We will investigate the behavior of our test statistic under the null and under the alternatives for a generic estimate of the slope with suitable rate of convergence. Next, we will get into the details in the standard case of the slope estimate based on the functional principal component analysis. In particular, we will see how the difficulty of estimating the parameters in the functional regression model could perturb the properties of the test. To keep the technical conditions readable, hereafter we will assume that $\mathbb{E}(|U|^m) < \infty$ for any $m \geq 1$.

3.4.1 The test statistic and the behavior under the null hypothesis

The test statistic is similar to the one we proposed for testing the effect of a functional covariate. Let β_0, $\gamma_0^{(p)}$, $\mathcal{S}^p$ and B_p be defined as in section 3.3. Let $\widehat{b} \in L^2[0,1]$ denote a generic estimator of the slope b and let

$$\widehat{a} = \overline{Y}_n - \int_0^1 \widehat{b}(t)\overline{X}_n(t)dt = a - \int_0^1 \{\widehat{b}(t) - b(t)\}\overline{X}_n(t)dt + \overline{U}_n,$$

where $\overline{U}_n = n^{-1}\sum_{i=1}^n U_i$. Let $\widehat{U}_i = Y_i - \widehat{a} - \langle \widehat{b}, X_i\rangle$ be the residuals and let

$$Q_n(\gamma;\widehat{a},\widehat{b}) = \frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} \widehat{U}_i\widehat{U}_j \frac{1}{h} K_h\left(\langle X_i - X_j, \gamma\rangle\right), \quad \gamma \in \mathcal{S}^p,$$

where recall $K(\cdot)$ is a kernel, h a bandwidth and $K_h(\cdot) = K(\cdot/h)$. Let $\widehat{v}_n^2(\cdot;\widehat{a},\widehat{b})$ be an estimate of the variance of $nh^{1/2}Q_n(\cdot;\widehat{a},\widehat{b})$ like in section 3.3.2. Given $B_p \subset \mathcal{S}^p$ with strictly positive Lebesgue measure in $\mathcal{S}^p$ that contains $\gamma_0^{(p)}$, the least favorable direction γ for H_0 is defined as

$$\widehat{\gamma}_n = \arg\max_{\gamma\in B_p}\left[nh^{1/2}Q_n(\gamma;\widehat{a},\widehat{b})/\widehat{v}_n(\gamma;\widehat{a},\widehat{b}) - \alpha_n \mathbb{I}_{\left\{\gamma\neq\gamma_0^{(p)}\right\}}\right]. \tag{3.4.12}$$

The test statistic is then

$$T_n = nh^{1/2}\frac{Q_n(\widehat{\gamma}_n;\widehat{a},\widehat{b})}{\widehat{v}_n(\widehat{\gamma}_n;\widehat{a},\widehat{b})}. \tag{3.4.13}$$

We will show that an asymptotic α-level test is given by $\mathbb{I}(T_n \geq z_{1-a})$, where z_a is the $(1-a)-$quantile of the standard normal distribution.

To derive the standard normal behavior of the test statistic under the null, we will show that under suitable conditions

$$\sup_{\gamma\in\mathcal{S}^p} nh^{1/2}|Q_n(\gamma;\widehat{a},\widehat{b}) - Q_n(\gamma)| = o_{\mathbb{P}}(1) \quad \text{and} \quad \sup_{\gamma\in\mathcal{S}^p} |\widehat{v}_n(\gamma;\widehat{a},\widehat{b})/\widehat{v}_n(\gamma) - 1| = o_{\mathbb{P}}(1), \tag{3.4.14}$$

with $Q_n(\gamma)$ and $\widehat{v}_n(\gamma)$ defined like in section 3.3, that is we will bring the problem back to the case where the errors U_i are observed.

Lemma 4.1 *Assume the conditions of Theorem 3.3 are met, $\mathbb{E}(|U|^m) < \infty$ for any $m \geq 1$, and $\int_0^1 \mathbb{E}[X^2(t)]dt < \infty$. Let $\widehat{b} \in L^2[0,1]$ be an estimator of b such that $\|\widehat{b} - b\|_{L^2} = O_{\mathbb{P}}(n^{-\rho})$ for some $3/8 < \rho \leq 1/2$. Moreover, suppose that the bandwidth h is such that $n^{1-2\zeta}h^{1/2} \to 0$ for some $3/8 < \zeta < \rho$. Then, under the hypothesis H_0 the uniform rates of convergence in (3.4.14) holds true.*

At this stage it is worthwhile to notice an important difference between the functional data framework and the finite-dimension case. In the later case the parameters of a parametric regression model could be easily estimated at the usual rate $O_{\mathbb{P}}(n^{-1/2})$ which makes the equivalences (3.4.14) hold without any further conditions on the model. In the functional covariate and functional parameter case, the rate of $\|\widehat{b}-b\|$ depends on the regularities of the covariate and of the slope parameter and is in general less than $O_{\mathbb{P}}(n^{-1/2})$, see Hall and Horowitz (2007), Crambes, Kneip and Sarda (2009). To make the differences $\widehat{U}_i - U_i$ sufficiently small and hence to preserve the standard normal critical values for T_n defined in (3.4.13) one has to pay a price on the bandwidth h : slower rates of $\|\widehat{b}-b\|_{L^2}$ will require faster decreases for h, and this will result in a loss of power against sequences of local alternatives. Below, we will investigate these aspects in more detail in the case where the slope is estimated using the functional principal component approach.

Theorem 4.2 *Under the conditions of Lemma 4.1 and if the hypothesis H_0 holds true, the law of the test statistic T_n is asymptotically standard normal. Consequently the test given by $\mathbb{I}(T_n \geq z_{1-a})$, where z_a is the $(1-a)-$quantile of the standard normal distribution, has asymptotic level a.*

The proof of this theorem is a direct consequence of Lemma 4.1 and the arguments we used for Theorem 3.3, therefore we will omit it.

There are several possibilities to estimate the parameters of a functional linear model. Let us investigate our test in the case where the estimate $\widehat{b}$ is obtained using the functional principal component analysis (PCA) approach, which is a standard approach for estimating the slope b ; see for instance Ramsay and Silverman (2005) and Hall and Horowitz (2007). For the sake of completeness, let us briefly recall this estimation procedure. Let $\mathcal{K}(u,v) = \mathrm{Cov}(X(u), X(v))$, $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$ and

$$\widehat{\mathcal{K}}(u,v) = \sum_{i=1}^n (X_i(u) - \overline{X}_n(u))(X_i(v) - \overline{X}_n(v)).$$

Write the spectral expansions of $\mathcal{K}$ and $\widehat{\mathcal{K}}$ as

$$\mathcal{K}(u,v) = \sum_{j=1}^\infty \theta_j \phi_j(u)\phi_j(v), \qquad \widehat{\mathcal{K}}(u,v) = \sum_{j=1}^\infty \widehat{\theta}_j \widehat{\phi}_j(u)\widehat{\phi}_j(v),$$

where

$$\theta_1 > \theta_2 > \cdots > 0, \qquad \widehat{\theta}_1 \geq \widehat{\theta}_2 \geq \cdots \geq 0$$

are the eigenvalues sequences of the operators with kernel $\mathcal{K}$ and $\widehat{\mathcal{K}}$, respectively, and $\phi_1, \phi_2, \ldots$ and $\widehat{\phi}_1, \widehat{\phi}_2, \ldots$ are the respective orthonormal eigenfunctions sequences. The linear operator corresponding to $\mathcal{K}$ is defined by $(\mathcal{K}f)(u) =$

$\int \mathcal{K}(u,v)f(v)\mathrm{d}v$. We have $\mathcal{K}b = g$ where $g(u) = \mathbb{E}[(Y - \mathbb{E}(Y))(X(u) - \mathbb{E}(X(u))]$. This suggests the estimator

$$\widehat{b}(t) = \sum_{j=1}^{m} \widehat{b}_j \widehat{\phi}_j(t), \quad t \in [0,1], \tag{3.4.15}$$

where the truncation point m is a smoothing parameter, $\widehat{b}_j = \widehat{\theta}_j^{-1}\widehat{g}_j$, $\widehat{g}_j = \langle \widehat{g}, \widehat{\phi}_j \rangle$ and

$$\widehat{g}(t) = n^{-1} \sum_{i=1}^{n} (Y_i - \overline{Y}_n)(X_i(t) - \overline{X}_n(t)) \tag{3.4.16}$$

with $\overline{Y}_n = n^{-1} \sum_{i=1}^{n} Y_i$.

For simplicity, hereafter the orthonormal basis $\mathcal{R}$ introduced in section 3.2 is the basis composed of the sequence of orthonormal eigenfunctions $\phi_1, \phi_2, \ldots$ of the covariance operator $\mathcal{K}$. Hence $X(t) = \sum_{j=1}^{\infty} x_j \phi_j(t)$, where the random coefficients $x_j = \langle X, \phi_j \rangle$. The following assumptions are standard conditions on the covariance operator $\mathcal{K}$ and the slope parameter in the linear model, as could be found in Hall and Horowitz (2007).

Assumption P

(a) The covariate X has finite fourth moment, that is $\int_0^1 \mathbb{E}[X^4(t)]dt < \infty$; moreover for some constant $C > 1$, $\mathbb{E}[x_j - \mathbb{E}(x_j)]^4 \leq C\theta_j^2$ for all j.

(b) The errors U_i are identically distributed, independent of X_i, with zero mean and finite variance.

(c) The eigenvalues θ_j of the covariance operator $\mathcal{K}$ satisfy

$$\theta_j - \theta_{j+1} \geq C^{-1} j^{-\alpha-1} \qquad \forall j \geq 1.$$

(d) The Fourier coefficients b_j satisfy

$$|b_j| \leq C j^{-\beta}$$

and $\alpha > 1$, $\frac{3}{2}\alpha + 2 < \beta$.

The condition $\frac{3}{2}\alpha + 2 < \beta$ replaces condition $\frac{1}{2}\alpha + 1 < \beta$ of Hall and Horowitz (2007) in order to conciliate the various requirements on the bandwidth h. For comparison, see also Theorem 4.1 of Cai and Hall (2006) where is required $\beta \geq \alpha + 2$. Hall and Horowitz (2007) show that if Assumption P holds true and if $m \asymp n^{1/(\alpha+2\beta)}$, then

$$\|\widehat{b} - b\|_{L^2}^2 = O_{\mathbb{P}}\left(n^{-\frac{2\beta-1}{\alpha+2\beta}}\right),$$

and this rate is optimal in a minimax sense. In this case $\rho = (2\beta - 1)/\{2(\alpha + 2\beta)\}$ and the condition $\rho > 3/8$ of Theorem 3.3, which guarantees a non empty range for the bandwidth, becomes $\beta > \frac{3}{2}\alpha + 2$, that is Assumption P-(d).

3.4.2 The behavior under the alternatives

The alternatives of the functional linear model we consider are of the form

$$H_{1n}: \; Y_{in} = a + \langle b, X_i\rangle + r_n\delta(X_i) + U_i^0, \quad n \geq 1, \quad \text{with } \mathbb{E}(U_i^0 \mid X_i) = 0, \quad 1 \leq i \leq n, \tag{3.4.17}$$

with $\delta(\cdot)$ a real-valued function such that $0 < \mathbb{E}[\delta^4(X)] < \infty$ and r_n, $n \geq 1$ a sequence of real numbers.

To be able to investigate the behavior of the test statistic under the alternatives, first we have to analyze the behavior of $\widehat{b}$, the estimator of b. To keep the paper at a reasonable length, hereafter we consider that $\widehat{b}$ is the estimator obtained through that functional PCA approach. In Lemma 4.3 below we derive the rate of convergence of $\widehat{b}$ towards b under the alternatives H_{1n}, provided that the function $\delta(\cdot)$ satisfies the orthogonality conditions

$$\mathbb{E}[\delta(X)] = 0 \qquad \text{and} \qquad \mathbb{E}[\delta(X)X] = 0. \tag{3.4.18}$$

Such orthogonality conditions are quite common in nonparametric testing, see for instance equation (3.11) in Guerre and Lavergne (2005), and they allow to focus on the performance of the test to detect departures from the model.

Lemma 4.3 *Assume that $X_1, \ldots, X_n$ are independent draws from X, $\int_0^1 \mathbb{E}[X^2(t)]dt < \infty$ and condition (3.4.18) holds true. Let $\widehat{b}$ (resp. $\widehat{b}^0$) be the estimator defined in (3.4.15) obtained from data generated according to the model (3.4.17) with a bounded sequence $r_n \geq 0$, $n \geq 1$ (resp. with $r_n = 0$ for all $n \geq 1$). Then*

$$\|\widehat{b}^0 - \widehat{b}\|_{L^2}^2 = O_{\mathbb{P}}(r_n^2 n^{-1}) \sum_{j=1}^{m} \widehat{\theta}_j^{-2}.$$

If in addition assumption P hold true and $m \asymp n^{1/(\alpha+2\beta)}$, then

$$\int_0^1 \{\widehat{b}(t) - b(t)\}^2 dt = O_{\mathbb{P}}\left(n^{-\frac{2\beta-1}{\alpha+2\beta}}\right) + o_{\mathbb{P}}(r_n^2).$$

Let us note that no moment condition for U^0 is needed for the proof of the first part of Lemma 4.3. Moreover, let us point out that we will not need to investigate the convergence rate for the estimator of a under the alternatives since by construction

$$\widehat{a} - a = -\int_0^1 \{\widehat{b}(t) - b(t)\}\overline{X}_n(t)dt + \overline{U}_n.$$

Now, we can analyze the behavior of the test statistics under the alternatives (3.4.17). The estimated residuals $\widehat{U}_i$ can be decomposed

$$\widehat{U}_i = U_i^0 + r_n\delta(X_i) - \langle \widehat{b} - b, X_i - \overline{X}_n\rangle - r_n\overline{\delta(X)}_n - \overline{U^0}_n \tag{3.4.19}$$

Theorem 4.4 *Consider the sequence of alternative hypotheses (3.4.17) with a nonzero function δ satisfying (3.4.18) and $0 < \mathbb{E}[\delta^4(X)] < \infty$. Let $\widehat{b} \in L^2[0,1]$ be an estimator of the slope parameter b. Suppose that the conditions of Theorem 3.3 are met with U replaced by U^0. Moreover, assume that $\widehat{b}$, the sequence r_n, $n \geq 1$, the sequence α_n, $n \geq 1$ and the bandwidth h satisfy the additional conditions :*

(i) $r_n^2 nh^{1/2}/\alpha_n \to \infty$;

(ii) $r_n^{-1}\|\widehat{b} - b\|_{L^2} = o_{\mathbb{P}}(1)$;

(iii) $\alpha_n/\{p^{3/2}\ln n\} \to \infty$.

Then the test based on T_n defined in (3.4.13) will reject the functional linear regression model with probability tending to 1, provided there exists $p \geq 1$ and $\widetilde{\gamma} \in B_p$ such that at least one of conditions (1) and (2) of Theorem 3.4 holds true.

If Assumption P holds true, condition (ii) of Theorem 4.4 indicates that the test could detect only local alternatives H_{1n} that approach the null hypothesis slower than $n^{-(2\beta-1)/\{2(\alpha+2\beta)\}}$. Meanwhile, in order to detect the fastest possible alternatives, the bandwidth should decrease to zero as slow as allowed by condition (i), that is faster than $n^{-2(\alpha+1)/(\alpha+2\beta)}$ times a power of $\ln n$, provided the dimension p and α_n increase as fast as a power of $\ln n$ such that condition (iii) is met.

3.5 Empirical analysis

3.5.1 Bootstrap procedures

To improve the critical values of the test statistic T_n with small samples we consider two bootstrap procedures that can be applied in both cases we consider herein : the U_i's are observed or the U's are estimated by some $\widehat{U}_i$'s. A bootstrap sample is denoted by U_i^b, $1 \leq i \leq n$, or $\widehat{U}_i^b$, $1 \leq i \leq n$. The first resampling procedure is a wild bootstrap procedure like in Mammen (1993), see also Li and Wang (1998) : $U_i^b = Z_iU_i$ (resp. $\widehat{U}_i^b = Z_i\widehat{U}_i$), $1 \leq i \leq n$, with $Z_i = V_i/\sqrt{2} + (V_i^2 - 1)/2$ and V_i independent standard normal variables independent from the original observations. The second resampling method we consider is a version of the smooth conditional moments bootstrap introduced by Gozalo (1997). If one looks at the statistic $nh^{1/2}Q_n(\gamma_0^{(p)})/\widehat{v}_n(\gamma_0^{(p)})$ which eventually provides the asymptotic critical values, one can notice that this quantity is exactly a test statistic for testing the moment condition $\mathbb{E}[U \mid \langle X, \gamma_0^{(p)} \rangle] = 0$. Therefore, following Gozalo (1997) and Lavergne and Patilea (2008), consider $U_i^b = \widehat{U}_i^b = \widehat{\sigma}_1(\langle X, \gamma_0^{(p)} \rangle) Z_i$, $1 \leq i \leq n$, where $\sigma_1(t)$ is some nonparametric estimate of $\sigma_1(t) = Var(U \mid \langle X, \gamma_0^{(p)} \rangle = t)$ satisfying (3.3.8). A bootstrap test statistic is built from a bootstrap sample as was the original test statistic. When this scheme is repeated many times, the bootstrap critical value $z^\star_{1-\alpha,n}$ at level α is the empirical $(1-\alpha)-$th quantile of the bootstrapped test statistics. This critical value is then compared to the initial test statistic.

3.5.2 Simulations for the test of effect

We used 1000 samples of $(U_1, X_1), \ldots, (U_n, X_n)$ of sizes $n = 100$ and $n = 200$, where X_i is a standard Brownian motion on the unit interval $[0, 1]$.

For the distribution of U_i, three scenarios were considered :

Null hypothesis $U_1, \ldots, U_n$ are i.i.d. $N(0, \sigma^2)$, and independent of $X_1, \ldots, X_n$, where $\sigma = 1.219$ (it is the same value used in the next scenario).

First alternative (Linear effect)

$$U_i = \langle b, X_i \rangle + U_i^0$$

where $b(t) = (\sin(2\pi t^3))^3$, and $U_1^0, \ldots, U_n^0$ are i.i.d. $N(0, \sigma^2)$, where $\sigma = 1.219$, corresponding to a 10 percent signal-to-noise ratio, in the sense of Cardot et al. (2003), that is, $E(\langle b, X \rangle^2)/(E(\langle b, X \rangle^2) + \sigma^2) = 0.1$.

Second alternative (Quadratic effect)

$$U_i = \int_0^1 \int_0^1 h(s,t) X(s) X(t)\, ds\, dt + U_i^0$$

where $h(s,t) = 0.6$, and $U_1^0, \ldots, U_n^0$ are i.i.d. $N(0, \sigma^2)$, where $\sigma = 1$.

Let us recall that the Karhunen-Loève expansion of the Brownian motion X, is given by

$$X(t) = \sum_{j=1}^{\infty} x_j \, \frac{1}{(j-0.5)\pi} \, \sqrt{2} \sin\left((j-0.5)\pi t\right)$$

where x_j are independent standard normal coefficients, $\mathcal{R} = \{\rho_j(t) = \sqrt{2}\sin((j - 0.5)\pi t) : j \in \{1, 2, \ldots\}\}$ constitutes an orthonormal basis of eigenfunctions, and $1/((j-0.5)^2\pi^2)$ are eigenvalues.

We will make use of this basis $\mathcal{R}$ and consider two possible values for p, $p = 3$ and $p = 5$, where p is the number of basic elements taken for the projections.

We only offer the results for this basis. The results obtained with other basis were quite similar. The role played by the basis and the dimension p consists of allowing to approximate both the covariate function X and the alternative. A good basis is that which provides a good approximation with a small dimension p. The Karhunen-Loève basis is obviously a good basis to approximate the covariate function. We will see the effect of the approximation of the alternative in the comparison between dimensions $p = 3$ and $p = 5$.

One major goal of this practical study, even more important than the basis, was to evaluate to effect of the direction $\gamma_0^{(p)}$ and the penalization α_n. Under the linear alternative, a good choice for $\gamma_0(p)$ is the projection of b on the basis, a bad choice would be an orthogonal direction in the same basis, and an uninformative direction could consist of giving the same weight to each basic element. As a bad choice we considered the projection of the function $\sin(10\pi t)$ on the basis.

Under the quadratic alternative, it is not so clear how to find a good direction $\gamma_0^{(p)}$. In this case we simply took the first eigenfunction in the Karhunen-Loève basis as a possibly good direction and the second eigenfunction as a possibly bad direction.

Different values for the penalization were considered. Since the statistic is standardized before penalization, natural values for α_n are 3, 4, 5 or 6. Small values of the penalization provide results that are similar to those obtained with the direction maximizing the standardized statistic, that is, $\arg\max_{\gamma\in\mathcal{S}^p} nh^{1/2}\frac{Q_n(\gamma)}{\hat{v}_n(\gamma)}$, while larger values of the penalization lead to results similar to those obtained with the chosen direction $\gamma_0^{(p)}$. The results presented here correspond to the penalization $\alpha_n = 5$.

To compute the statistic for each direction, we used the Epanechnikov kernel, $K(x) = 1 - x^2$ for $x \in [-1, 1]$, and we selected the bandwidth as $h = c_h n^{-2/9}$, with three values for the constant $c_h = 0.6, 0.8, 1.0$.

To estimate the conditional variance, the two estimators (3.3.5) and (3.3.7) were considered. For the estimator (3.3.7), a kernel estimator of the errors' conditional variance was used, with uniform kernel and bandwidth $h_v = 0.5n^{-1/6}$.

For the optimization in the hypersphere $\mathcal{S}^p$, a grid of 300 points was used in the case of $p = 3$ dimensions, and a grid of 1280 points in the case of $p = 5$ dimensions. Aditionally, a local refinement of the optimum was used, with 9 points in dimension $p = 3$ and 81 points in dimension $p = 5$. For each original sample, we used 199 bootstrap samples to compute the critical value.

Table 1 below contains the percentages of rejections under the null hypothesis, for the best direction (projections of b on the basis), while Table 2 contains the same type of results for the worst direction (orthogonal to the best direction in the basis). These directions are best and worst under the linear alternative, but

		$p=3$				$p=5$			
		$c_h=0.8$	$c_h=1.0$	$c_h=1.2$	CT	$c_h=0.8$	$c_h=1.0$	$c_h=1.2$	CT
$n=100$	(3.5)	5.0	4.9	4.9	6.6	5.4	4.9	5.4	6.7
	(3.7)	6.5	5.7	5.8	6.6	7.1	6.4	6.0	6.7
$n=200$	(3.5)	4.3	4.5	4.8	4.8	4.0	4.7	4.6	5.4
	(3.7)	5.4	5.5	5.5	4.8	5.2	5.6	5.5	5.4

TABLE 3.1 – Percentage of rejections under the null hypothesis, with nominal level 5% (best direction).

		$p=3$				$p=5$			
		$c_h=0.8$	$c_h=1.0$	$c_h=1.2$	CT	$c_h=0.8$	$c_h=1.0$	$c_h=1.2$	CT
$n=100$	(3.5)	5.1	4.8	4.8	6.6	4.6	4.8	4.6	6.7
	(3.7)	6.6	6.1	5.9	6.6	7.7	5.9	5.7	6.7
$n=200$	(3.5)	4.8	5.0	5.1	4.8	4.3	4.7	4.7	5.4
	(3.7)	6.2	6.1	5.9	4.8	5.5	5.3	5.6	5.4

TABLE 3.2 – Percentage of rejections under the null hypothesis, with nominal level 5% (worst direction).

not under the null. These results are shown to prove that the level is correct under different chosen directions.

Results are provided with different values of the bandwidth (represented by the constant c_h), and with the conditional variance estimated by (3.3.5) and (3.3.7), as indicated in the title of the rows. The columns titled CT contain the percentages corresponding to Cardot et al. (2003) test.

Tables 1 and 2 show a good behaviour of our test under the null hypothesis, since the percentages of rejections are generally close to the nominal level, 5%.

Tables 3 and 4 below contain the percentages of rejection under the linear alternative, with the best direction for Table 3 and the worst direction for Table 4.

Cardot et al. (2003) test is clearly more powerful than our test, which was to be expected, because their test is designed to detect the linear effect, under the linear model. Meanwhile, our test is consistent under any alternative, including nonlinear

		$p=3$				$p=5$			
		$c_h=0.8$	$c_h=1.0$	$c_h=1.2$	CT	$c_h=0.8$	$c_h=1.0$	$c_h=1.2$	CT
$n=100$	(3.5)	49.6	48.6	47.3	79.1	50.0	49.3	47.5	72.4
	(3.7)	50.9	52.3	50.2	79.1	50.1	52.8	51.0	72.4
$n=200$	(3.5)	83.7	84.3	83.4	98.0	81.6	84.0	84.6	96.5
	(3.7)	85.7	85.9	86.2	98.0	82.1	85.6	86.2	96.5

TABLE 3.3 – Percentage of rejections under the linear alternative, with nominal level 5% (best direction).

ones.

Anyway the power shown by our test is good in most cases, with similar results in the reasonably wide range of bandwidths considered here. For a bigger dimension, from $p=3$ to $p=5$, our test provides similar results while Cardot test has less power. The reason is that the increase of dimension produces a better approximation of the function b in the basis, but at the same time the increased dimension gives place to more noise in the statistic. These two contradictory effects are balanced for our test, while Cardot test is more affected by the noise coming from a bigger dimension.

The powers obtained with the variance estimation given by (3.3.7) are higher than those obtained with the variance estimation given by (3.3.5). This was expected because the variance estimator given by (3.3.7) provides smaller denominators for the standardized statistic under the alternative. This way the usefulnees of the more complicated variance estimation given by (3.3.7) is justified.

For the worst direction (Table 4) the power is substantially lower than for the best direction (Table 3). This was expected. In fact, the test based only on the worst direction, without the contribution of the maximization in other directions, has no power at all. The conclusion is that even a very bad choice of the privileged direction can be compensated by the maximization, in order to provide a consistent test, even though the power will not be the same as with the best direction. At the end, most of the chosen directions (uninformative ones, for example) will provide a reasonable power. The amount of penalization will depend on the certainty about the possible direction of the alternative, so a good direction would be privileged with a bigger penalization.

Tables 5 and 6 contain the results under the quadratic alternative. As it is shown, Cardot test is not consistent at all under this alternative, which was expected because it is only designed to detect linear effects.

		$p = 3$				$p = 5$			
		$c_h = 0.8$	$c_h = 1.0$	$c_h = 1.2$	CT	$c_h = 0.8$	$c_h = 1.0$	$c_h = 1.2$	CT
$n = 100$	(3.5)	23.0	20.4	18.6	79.1	28.3	23.5	18.6	72.4
	(3.7)	31.9	29.5	25.0	79.1	38.6	33.3	27.8	72.4
$n = 200$	(3.5)	53.2	51.6	49.4	98.0	64.7	60.4	55.2	96.5
	(3.7)	59.9	58.8	56.3	98.0	70.8	66.4	62.3	96.5

TABLE 3.4 – Percentage of rejections under the linear alternative, with nominal level 5% (worst direction).

		$p = 3$				$p = 5$			
		$c_h = 0.8$	$c_h = 1.0$	$c_h = 1.2$	CT	$c_h = 0.8$	$c_h = 1.0$	$c_h = 1.2$	CT
$n = 100$	(3.5)	20.3	22.2	23.0	8.6	19.3	22.2	23.0	8.1
	(3.7)	24.1	26.0	27.7	8.6	23.4	25.6	27.5	8.1
$n = 200$	(3.5)	35.4	39.7	43.1	7.6	33.7	38.9	42.4	6.9
	(3.7)	41.6	44.3	47.6	7.6	39.1	43.4	47.0	6.9

TABLE 3.5 – Percentage of rejections under the quadratic alternative, with nominal level 5% (first eigenfunction as chosen direction).

In this situation there is no clear reference of a good privileged direction. Table 5 contains the results corresponding to the first eigenfunction taken as the chosen direction, while Table 6 corresponds to the second eigenfunction as chosen direction.

The conclusions regarding the variance estimation are similar to those obtained under the linear alternative, with more power for the variance estimation given by (3.3.7).

Regarding the comparison of chosen directions and dimensions $p = 3$ and $p = 5$, we can conclude that the first eigenfunction (Table 5) with the smaller dimension $p = 3$ is the combination with highest power. To obtain a similar power with the second eigenfunction, one should increase the dimension from $p = 3$ to $p = 5$, so other directions have more influence on the result.

		$p=3$				$p=5$			
		$c_h=0.8$	$c_h=1.0$	$c_h=1.2$	CT	$c_h=0.8$	$c_h=1.0$	$c_h=1.2$	CT
$n=100$	(3.5)	11.4	11.1	10.4	8.6	16.2	13.4	12.1	8.1
	(3.7)	18.2	16.6	14.9	8.6	23.8	20.6	17.1	8.1
$n=200$	(3.5)	24.1	23.7	22.6	7.6	36.9	33.2	29.5	6.9
	(3.7)	29.3	28.8	27.0	7.6	41.5	38.4	34.4	6.9

TABLE 3.6 – Percentage of rejections under the quadratic alternative, with nominal level 5% (second eigenfunction as chosen direction).

3.5.3 Simulations for the goodness-of-fit test

Now we will present simulation results for the goodness-of-fit test. First, the goodness-of-fit test of the functional linear model, as studied in Section 4, will be analyzed under the null and under different alternatives. Later, the goodness-of-fit of a functional quadratic model will also be studied to show the applicability of our method to general functional parametric models.

The functional linear model, as considered in Section 4, is given by

$$Y = a + \langle b, X\rangle + U$$

Here X is drawn as a Brownian motion, $b \in L^2[0,1]$ and $a \in \mathbb{R}$ are parameters to be estimated, and $U = \delta(X) + U^0$, where $\delta(X)$ is the deviation from the null hypothesis, and U^0 is the error. For the parameters, $b(t) = 1$ for all $t \in [0,1]$ and $a = 0$ were taken as the true values.

A sample $(Y_1, X_1), \dots (Y_n, X_n)$ will be drawn from this model (of sizes $n = 100$ and $n = 200$), that is,

$$Y_i = a + \langle b, X_i\rangle + \delta(X_i) + U_i^0 \qquad \text{for} \quad i \in \{1, \dots, n\}$$

where $U_1^0, \dots, U_n^0$ are independent standard normal variables, also independent of $X_1, \dots, X_n$.

The function $\delta(\cdot)$ is taken to be zero under the null hypothesis, and different from zero under the alternative. We consider one scenario under the null and two different scenarios under the alternative :

Null hypothesis

$$\delta(X_i) = 0$$

First alternative (Quadratic deviation)

$$\delta_2(X_i) = 0.6 \left(\int_0^1 \int_0^1 X(s)X(t)\, ds\, dt - 1/3 \right)$$

Second alternative (Cubic deviation)

$$\delta_3(X_i) = 0.9 \left(\int_0^1 \int_0^1 \int_0^1 X(s)X(t)X(z)\, ds\, dt\, dz - \int_0^1 X(t)dt \right)$$

Note that the two functions $\delta_2(\cdot)$ and $\delta_3(\cdot)$ satisfy the orthogonality conditions (4.18).

Tables 7, 8 and 9 below contain, respectively, the percentages of rejection under the null hypothesis and under the quadratic and cubic alternatives. Since the alternatives are non-linear, symmetric and orthogonal to linear effects, there is no clear reference of a better privileged direction. For this reason, results are only given for an uninformative direction (with the same coefficients in all basic elements) and with penalization $\alpha_n = 5$.

Since the previous results on the test of effect have shown a better behaviour of the variance estimator (3.3.7), results will only be given for this type of variance estimator.

Now, a new parameter has to be decided, which is the dimension m in the estimation of the functional linear model (see Section 4). We used values $m = 1$ and $m = 3$.

The Karhunen-Loève expansion of the Brownian motion X is again used as the basis with $p = 3$ and $p = 5$ as the number of basic functions. The kernels and bandwidths for computing the test statistic and the variance estimator were chosen with the same criteria as in the case of the test of effect.

Recently, Horvath and Reeder (2011) have proposed a test of significance of the quadratic effect under a functional quadratic model. For purposes of comparison, percentages of rejection of this test are included in Tables 7, 8 and 9. Note that Horvath and Reeder's test is specially designed to detect quadratic alternatives to the linear model, as the one proposed here as first alternative, whose results are shown in Table 8.

Results from Horvath and Reeder's test are given in the column titled "HR". Their test requires the choice of a dimension in a principal component decomposition of the covariate. Then, in Tables 7, 8 and 9, the value of m represents this dimension for Horvath and Reeder's test.

As shown in Table 7, the level is quite well respected by our test under the null hypothesis.

			$p = 3$			$p = 5$		
		HR	$c_h = 0.8$	$c_h = 1.0$	$c_h = 1.2$	$c_h = 0.8$	$c_h = 1.0$	$c_h = 1.2$
$n = 100$	$m = 1$	5.9	5.8	5.9	5.1	5.8	5.5	5.1
	$m = 3$	11.0	5.6	5.3	5.7	5.2	5.0	5.0
$n = 200$	$m = 1$	6.3	6.2	5.9	5.7	6.0	6.5	6.4
	$m = 3$	6.4	5.6	4.8	4.9	5.3	5.0	5.1

TABLE 3.7 – Percentage of rejections under the null hypothesis (linear model), with nominal level 5%.

In Table 8, higher power is obtained by Horvath and Reeder's test for detecting the quadratic alternative. This was expected again. However, their test is also inconsistent under general alternatives, as shown in Table 9, where a cubic alternative is studied.

As a general conclusion, the test proposed here is consistent under general alternatives and provides reasonable power even in the comparison with parametric tests.

Regarding the parameter m, it plays the role of a regularization parameter in the estimation of the functional linear model both in the Horvath and Reeder's test and in our test. However, for the Horvath and Reeder's test m is also used to estimate the quadratic effects, while a different parameter, namely p, is the number of basic elements used to represent the deviation in our test.

An increased value of m would yield a less biased estimation of the linear effects (with constant coefficients $b(t) = 1$ for all $t \in [0, 1]$), at the expense of a bigger variance and a more noisy test statistic. The observed results show that a bigger value of m produces less power for the Horvath and Reeder's test while our test is not very affected or even improved under the cubic alternative.

An increased value of p provides better power for our test, as a consequence of a better approximation of the quadratic and cubic deviations (with constant kernels) in the basis.

As an illustration of the behaviour of our test for the goodness-of-fit of a more general parametric model, we considered the goodness-of-fit of the quadratic functional model, and obtained percentages of rejection under the null hypothesis and under the cubic alternative. That is, the simulated model would be

$$Y = a + \int_0^1 b(t)X(t)\,dt + \int_0^1 \int_0^1 h(s,t)X(s)X(t)\,ds\,dt + \delta(X) + U^0$$

			$p=3$			$p=5$		
		HR	$c_h=0.8$	$c_h=1.0$	$c_h=1.2$	$c_h=0.8$	$c_h=1.0$	$c_h=1.2$
$n=100$	$m=1$	78.3	29.0	32.9	35.8	31.3	35.8	38.6
	$m=3$	60.8	27.0	30.9	34.3	29.6	32.1	36.6
$n=200$	$m=1$	96.1	54.8	59.9	64.0	58.5	64.3	69.3
	$m=3$	86.5	53.5	59.5	63.2	57.7	64.5	69.0

TABLE 3.8 – Percentage of rejections under the quadratic alternative to the linear model, with nominal level 5%.

			$p=3$			$p=5$		
		HR	$c_h=0.8$	$c_h=1.0$	$c_h=1.2$	$c_h=0.8$	$c_h=1.0$	$c_h=1.2$
$n=100$	$m=1$	29.2	36.7	38.6	38.6	35.9	34.9	34.5
	$m=3$	29.6	36.9	39.6	42.5	36.7	37.9	37.7
$n=200$	$m=1$	31.0	65.4	69.8	72.9	66.2	68.7	69.9
	$m=3$	28.9	69.8	73.2	76.0	69.4	72.0	75.0

TABLE 3.9 – Percentage of rejections under the cubic alternative to the linear model, with nominal level 5%.

		$p = 3$			$p = 5$		
		$c_h = 0.8$	$c_h = 1.0$	$c_h = 1.2$	$c_h = 0.8$	$c_h = 1.0$	$c_h = 1.2$
$n = 100$	$m = 1$	6.3	6.1	5.8	5.3	5.1	4.7
	$m = 3$	4.9	4.5	4.0	4.8	4.6	5.3
$n = 200$	$m = 1$	5.7	5.5	6.0	5.9	6.6	7.0
	$m = 3$	4.8	4.5	3.9	4.4	4.4	5.3

TABLE 3.10 – Percentage of rejections under the null hypothesis (quadratic model), with nominal level 5%.

		$p = 3$			$p = 5$		
		$c_h = 0.8$	$c_h = 1.0$	$c_h = 1.2$	$c_h = 0.8$	$c_h = 1.0$	$c_h = 1.2$
$n = 100$	$m = 1$	35.8	39.3	42.3	35.8	39.4	39.9
	$m = 3$	35.8	39.4	44.9	37.9	43.1	45.4
$n = 200$	$m = 1$	62.8	66.3	67.6	62.7	67.6	69.8
	$m = 3$	64.1	69.5	72.5	64.5	69.1	73.5

TABLE 3.11 – Percentage of rejections under the cubic alternative to the quadratic model, with nominal level 5%.

which consists of a quadratic functional model, as considered in Yao and Müller (2010) and Horvath and Reeder (2011), plus a deviation represented by the function $\delta(\cdot)$. Here $b(t) = 1$ for all $t \in [0, 1]$, and $h(s, t) = 0.6$ for all $s, t \in [0, 1]$, which are the same linear and quadratic effects considered before. The deviation was chosen to be $\delta = \delta_3$, that is, the cubic deviation already studied. Then, the idea will be to carry out a goodness-of-fit of the functional quadratic model, and evaluate its performance under the null $\delta = 0$ and under a cubic alternative $\delta = \delta_3$.

Results are given in Tables 10 and 11. To the best of our knowledge there is no parametric test for comparison in this situation, and then only the percentages of rejection of our test are included.

The results show a good behaviour under the null, where the level is generally respected. The power obtained under the alternative is as expected, with no much effect coming from the parameters m and p, a slightly bigger power for larger bandwidths, and of course increasing power as the sample size grows.

3.5.4 Application to real data

The test proposed here is applied to the data set collected by Tecator and available at http ://lib.stat.cmu.edu/datasets/tecator. The task is to predict the fat content of a meat sample on the basis of its near infrared absorbance spectrum. For each sample of finely chopped pure meat, a 100 channel spectrum of absorbances was recorded using a Tecator Infratec Food and Feed Analyzer, a spectrometer that works in the wavelength range 850-1050 nm. These absorbances can be thought of as a discrete approximation to the continuous record, $X_i(t)$. Also, for each sample of meat, the fat content, Y_i, was measured by analytic chemistry. The data set contains 240 samples of meat.

Yao and Müller (2010) proposed using a functional quadratic model to predict the fat content, Y_i, of a meat based on its absorbance spectrum, $X_i(t)$. Horvath and Reeder (2011) applied their parametric test to check whether the quadratic term is needed, versus the null hypothesis of a functional linear model. Their reached the conclusion that the quadratic effect is significant, and then, the functional quadratic model is needed.

We will apply the test proposed here, first to check the goodness-of-fit of the functional linear model, and later the goodness-of-fit of the functional quadratic model. Table 1 below contains the p-values corresponding to our test for different values of the bandwidth, the parameter m for model estimation and the dimension p. We can conclude that both the functional linear and the functional quadratic models should be rejected for the Tecator data set. This conclusion confirms the empirical results of Chen, Hall and Müller (2011) who proposed an additive double index model. Indeed, the link functions estimated by Chen, Hall and Müller do not show respective linear and quadratic patterns which indicates that the usual functional linear and the functional quadratic models do not provide a satisfactory fit.

		Linear model				Quadratic model			
	h	0.18	0.30	0.44	0.59	0.18	0.30	0.44	0.59
$p=2$	$m=1$	0.5	0.4	0.2	0.6	2.4	1.4	1.6	3.3
	$m=2$	0.2	0.0	0.0	0.3	0.6	0.3	0.0	0.7
	$m=3$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$p=3$	$m=1$	0.0	0.0	0.2	0.2	0.0	0.1	0.1	0.0
	$m=2$	0.0	0.0	0.0	0.1	0.2	0.0	0.1	0.0
	$m=3$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 1. p-values (in percentages) obtained by applying the new test to the Tecator data set.

3.6 Appendix

3.6.1 Dimension reduction : proof of the fundamental lemma

Proof of Lemma 2.1. (A). The implication (1) $\Rightarrow$ (2) is obvious. To prove (2) $\Rightarrow$ (1), note first that for any $\beta \neq 0$, the $\sigma-$field generated by $\langle X, \beta\rangle$ is the same as the $\sigma-$field generated by $\langle X, \beta\rangle / \|\beta\|_{L^2}$. Next, by elementary properties of the conditional expectation, we obtain that for any $\beta \in L^2[0,1]$, including $\beta = 0$,

$$\begin{aligned} 0 &= \mathbb{E}\left[\exp\{i\langle X, \beta\rangle\}\mathbb{E}(Z \mid \langle X, \beta\rangle)\right] \\ &= \mathbb{E}\left[\exp\{i\langle X, \beta\rangle\} Z\right] \\ &= \mathbb{E}\left[\exp\{i\langle X, \beta\rangle\}\mathbb{E}(Z \mid X)\right] . \end{aligned} \tag{3.6.1}$$

Write $Z = Z^+ - Z^-$ where Z^+ and Z^- are the positive and negative parts of Z, and deduce that for any β, $\mathbb{E}\left[\exp\{i\langle X, \beta\rangle\}\mathbb{E}(Z^+ \mid X)\right] = \mathbb{E}\left[\exp\{i\langle X, \beta\rangle\}\mathbb{E}(Z^- \mid X)\right]$. As distinct positive finite measures cannot have the same characteristic function, see for instance Theorem 3.1 of Parthasarathy (1967), this implies that $\mathbb{E}(Z^+ \mid X) = \mathbb{E}(Z^- \mid X)$ and hence $\mathbb{E}(Z \mid X) = 0$ almost surely. For (2) $\Rightarrow$ (3) it suffices to identify γ with an element in $L^2[0,1]$ of norm 1. To prove (3) $\Rightarrow$ (1), fix arbitrarily $\beta \in L^2[0,1]$, $\beta \neq 0$. For any $p \geq 1$, let $\beta^{(p)}$ be the projection of β on the subspace generated by the first p elements of the basis $\mathcal{R}$. For any p sufficiently large such that $\|\beta^{(p)}\| = \|\beta^{(p)}\|_{L^2} > 0$ we have $\mathbb{E}(Z \mid \langle X, \beta^{(p)}\rangle) = \mathbb{E}(Z \mid \langle X, \beta^{(p)} / \|\beta^{(p)}\|\rangle) = 0$, where for the last equality we use the fact that $\beta^{(p)}/\|\beta^{(p)}\| \in \mathcal{S}^p$ and (3). By elementary properties of the conditional expectation,

$$\begin{aligned} 0 &= \mathbb{E}\left[\exp\{i\langle X, \beta^{(p)}\rangle\}\mathbb{E}(Z \mid \langle X, \beta^{(p)}\rangle)\right] \\ &= \mathbb{E}\left[\exp\{i\langle X, \beta\rangle\} Z \exp\{i\langle X, \beta^{(p)} - \beta\rangle\}\right] \\ &= \mathbb{E}\left[\exp\{i\langle X, \beta\rangle\}\mathbb{E}(Z \mid X)[\exp\{i\langle X, \beta^{(p)} - \beta\rangle\} - 1]\right] \\ &\quad + \mathbb{E}\left[\exp\{i\langle X, \beta\rangle\}\mathbb{E}(Z \mid X)\right]. \end{aligned}$$

From the Taylor expansion with integral reminder and elementary calculus, one obtains that $\forall x \in \mathbb{R}$, $|\exp(ix) - 1| \leq \min\{|x|, 2\}$. From this and the Cauchy-Schwarz inequality, deduce that for any p,

$$|\exp\{i\langle X, \beta^{(p)} - \beta\rangle\} - 1| \leq \min\{\|X\|_{L^2}\|\beta^{(p)} - \beta\|_{L^2}, 2\}.$$

Since $\|\beta^{(p)} - \beta\|_{L^2} \to 0$ when $p \to \infty$, by Lebesgue Dominated Convergence Theorem it follows that necessarily

$$\mathbb{E}\left[\exp\{i\langle X, \beta\rangle\}\mathbb{E}(Z \mid X)\right] = 0.$$

Since $\beta \in L^2[0,1]$ was arbitrarily fixed, apply again the arguments we used after equation (3.6.1) to deduce that (1) hold true. The equivalence (3) $\Leftrightarrow$ (4) follows from Lemma 2.1-(A) of Lavergne and Patilea (2008) applied for each p.

(B). From (A)-4 above, there exists some $p_0 \geq 1$ such that $\mathbb{P}[\mathbb{E}(Z \mid X^{(p_0)}) = 0] < 1$. On the other hand, by the property of iterated expectations, for any $p > p_0$,

$$\mathbb{E}(Z \mid X^{(p_0)}) = \mathbb{E}[\mathbb{E}(Z \mid X^{(p)}) \mid X^{(p_0)}] \, .$$

Thus necessarily $\mathbb{P}[\mathbb{E}(Z \mid X^{(p)}) = 0] < 1, \forall p > p_0$. Fix arbitrarily $p > p_0$ and notice that for any $b \in [-1,1]$,

$$\{\gamma \in \mathcal{S}^p : \mathbb{E}(Z \mid \langle X, \gamma \rangle) = 0 \ a.s.\} \subset \{\gamma \in \mathcal{S}^p : \mathbb{E}(Z \exp\{b\langle X, \gamma\rangle\}) = 0\}.$$

The expectations in the sets in the last display are well-defined since

$$\mathbb{E}(|Z \exp\{b\langle X, \gamma\rangle\}|) \leq \mathbb{E}(|Z| \exp\{|b||\langle X, \gamma\rangle|\}) \leq \mathbb{E}(|Z| \exp\{\|X\|\}) < \infty.$$

Let us notice that

$$\{b\widetilde{\gamma} : b \in [-1,1],\ \gamma \in \mathcal{S}^p,\ \mathbb{E}(Z \exp\{b\langle X, \gamma\rangle\}) = 0\} \subset \widetilde{A}_p$$

where

$$\widetilde{A}_p := \{\widetilde{\gamma} \in \mathbb{R}^p : \|\widetilde{\gamma}\| \leq 1, \mathbb{E}(Z \exp\{\langle X^{(p)}, \widetilde{\gamma}\rangle\}) = 0\}.$$

Thus, to prove (B) it suffice to show that the set $\widetilde{A}_p$ has Lebesgue measure zero in $\mathbb{R}^{p-1}$ and is not dense in the unit ball of $\mathbb{R}^{p-1}$.[†]

For these purpose, we will use the following property : if W_1 and W_2 are real-valued random variables such that $\mathbb{E}(|W_1| \exp\{a|W_2|\}) < \infty$ for some $a > 1$, then

$$\mathbb{P}(\mathbb{E}(W_1 \mid W_2) = 0) < 1 \implies \text{ the set } \{|b| < a : \mathbb{E}(W_1 \exp\{bW_2\}) = 0\} \text{ is empty or finite.} \tag{3.6.2}$$

To prove this property, decompose $W_1 = W_1^+ - W_1^-$ and use the positive part W_1^+ to define

$$\lambda^+(b) = \mathbb{E}(W_1^+ \exp\{bW_2\}) = \mathbb{E}(\mathbb{E}(W_1^+ \mid W_2) \exp\{bW_2\}) = \int_{\mathbb{R}} \exp\{bw\} d\mu^+(w),$$

$|b| < a$, where $d\mu^+(w) = \mathbb{E}(W_1^+ \mid W_2 = w) dF_{W_2}(w)$ and F_{W_2} is the probability distribution function of W_2. Use the negative part of W_1 to define $\lambda^-(b)$, $|b| < a$ similarly. Since W_1 is integrable, $\mu^-(\mathbb{R}), \mu^+(\mathbb{R}) < \infty$. The functions $\lambda^-(\cdot)/\mu^-(\mathbb{R})$ and $\lambda^+(\cdot)/\mu^+(\mathbb{R})$ are the Laplace transforms of the probability distributions $\mu^-/\mu^-(\mathbb{R})$

†. An easy way to check that it is indeed sufficient to derive such properties for $\widetilde{A}_p$ is to represent the sets in the hyperspherical coordinates.

and $\mu^+/\mu^+(\mathbb{R})$. The condition $\mathbb{E}(|W_1|\exp\{a|W_2|\}) < \infty$ implies that these Laplace transforms, and hence $\lambda^+(\cdot)$ and $\lambda^-(\cdot)$, are (real) analytic on the domain $(-a, a)$. See for instance Proposition 8.4.4 in Chow and Teicher (1997). Notice that the set in (3.6.2) is the set of $b \in (-a, a)$ for which $\lambda^-(b) = \lambda^+(b)$. If $\mathbb{P}(\mathbb{E}(W_1 \mid W_2) = 0) < 1$, $\lambda^+(\cdot)$ and $\lambda^-(\cdot)$ cannot coincide on $(-a, a)$. Thus the set in (3.6.2) contains only isolated points b from the interval $(-a, a)$, which means that it is necessarily empty or finite.

Now, let fix some $1 < \zeta < s$. Recall that we want to investigate the cardinality of the $\widetilde{A}_p$, subset of the unit ball of $\mathbb{R}^p$. From (A), there exists $\widetilde{\gamma} \in \mathcal{S}^p$ such that $\mathbb{P}(\mathbb{E}(Z \mid \langle X^{(p)}, \widetilde{\gamma}\rangle) = 0) < 1$. Then, property (3.6.2) applied with some $a > 1$, $W_1 = Z$ and $W_2 = \langle X^{(p)}, \widetilde{\gamma}\rangle$ implies that the set

$$\{|b| < a : \mathbb{E}(Z\exp\{\langle X^{(p)}, b\widetilde{\gamma}\rangle\}) = 0\}$$

is empty or finite. Deduce that there exists $\upsilon^\star$ in the unit ball of $\mathbb{R}^p$, arbitrarily close to the origin, in particular with $\|\upsilon^\star\| < s - \zeta$, such that $\mathbb{E}(Z\exp\{\langle X^{(p)}, \upsilon^\star\rangle\}) \neq 0$. Next, we adapt the lines of the proof of Lemma 1 in Bierens (1990). Let $Z^\star = Z\exp\{\langle X^{(p)}, \upsilon^\star\rangle\}$. By construction, $\mathbb{P}(\mathbb{E}(Z^\star \mid x_1, \cdots, x_l) = 0) < 1$, for $l = 1, \cdots, p$. Define the sets

$$A_l^\star = \{(t_1, \cdots, t_l) \in \mathbb{R}^l : \|(t_1, \cdots, t_l)\| \leq \zeta, \mathbb{E}(Z^\star\exp\{(x_1t_1 + \cdots + x_lt_l)\}) = 0\},$$

$l = 1, \cdots, p$. Since $|t_1x_1 + \cdots + t_lx_l + \langle X^{(p)}, \upsilon^\star\rangle| \leq (\|\upsilon^\star| + \zeta\|X^{(p)}\|$, deduce from property (3.6.2) applied with $a = \zeta$, $W_1 = Z^\star$ and $W_2 = x_1$ that the set $A_1^\star$ is empty or finite. Now, define the set

$$A_2^{\star\star}(t_1) = \{|t_2| \leq (\zeta^2 - t_1^2)^{1/2} : \mathbb{E}(Z^\star\exp\{x_1t_1\}\exp\{x_2t_2\}) = 0\}.$$

If $|t_1| < \zeta$ but $t_1 \notin A_1^\star$, replace $Z^\star$ by $Z^\star\exp\{x_1t_1\}$ and use again property (3.6.2) with $a = (\zeta^2 - t_1^2)^{1/2}$, $W_1 = Z^\star\exp\{t_1x_1\}$ and $W_2 = x_2$ to deduce that the set $A_2^{\star\star}(t_1)$ is empty or finite. This means that $A_2^\star$ is contained in the union of some sets $B' \times \mathbb{R}$ and $\mathbb{R} \times B''$ where B' and B'' are empty or finite. Repeat the arguments with $l = 3, \cdots, p$ and deduce that $A_p^\star$ has Lebesgue measure zero in $\mathbb{R}^p$. Since the norm of $\upsilon^\star$ could be taken arbitrarily small such that $\widetilde{A}_p \subset A_p^\star$, we can now easily deduce that $\widetilde{A}_p$ has Lebesgue measure zero in the unit ball of $\mathbb{R}^p$. The fact that $\widetilde{A}_p$ is not dense in the unit ball of $\mathbb{R}^p$ is a direct consequence of the fact that $A_p^\star$ intersected with unit ball of $\mathbb{R}^p$ is not dense. ■

3.6.2 Rates of convergence : technical lemmas

For ν a probability measure on a sample space, $\mathcal{F}$ a class of functions and $\varepsilon > 0$, let $N(\varepsilon, \mathcal{F}, L^2(\nu))$, denote the covering number, that is the minimal number of balls of radius ε in $L^2(\nu)$ needed to cover $\mathcal{F}$. See Van der Vaart and Wellner (1996) or Kosorok (2008) for the definitions. For real random variables, $A_n \asymp_{\mathbb{P}} B_n$ means that there exists a constant $C > 1$ such that $\mathbb{P}(1/C \leq A_n/B_n \leq C)$ goes to 1 when n grows. In the following $C, C_1, c, c_1, \cdots$ represent constants that may change from line to line.

Lemma 6.1 *For any $p \geq 1$, let*

$$\mathcal{F}_{1p} = \{(v_1, v_2) \mapsto K(h^{-1}\langle v_1 - v_2, \gamma\rangle) : v_1, v_2 \in \mathbb{R}^p, \gamma \in \mathcal{S}^p, h > 0\}$$

and

$$\mathcal{F}_{2p} = \{v \mapsto \mathbb{E}[K(h^{-1}\langle X - v, \gamma\rangle) \,|: v \in \mathbb{R}^p, \gamma \in \mathcal{S}^p, h > 0\}.$$

If Assumption K-(a) holds, there exist constants $c_1, c_2, c_3 > 0$ such that for any $p \geq 1$ and $0 < \varepsilon < 1$ and any ν_1 probability measure on $\mathbb{R}^p \times \mathbb{R}^p$ and ν_2 probability measure on $\mathbb{R}^p$,

$$N(\varepsilon, \mathcal{F}_{jp}, L^2(\nu_j)) \leq c_1(c_2/\varepsilon)^{c_3 p}, \quad j = 1, 2. \tag{3.6.3}$$

Proof. Since K can be written as a difference of two monotone functions, the result for $\mathcal{F}_{1p}$ is an easy consequence of the Theorem 9.3, Lemmas 9.6 and 9.9 of Kosorok (2008) and Lemma 16 of Nolan and Pollard (1987) ; see also their Lemma 22-(ii). For $\mathcal{F}_{2p}$, use the bound for $\mathcal{F}_{1p}$ and Lemma 20 of Nolan and Pollard (1987). ■

Lemma 6.2 *Let Assumptions D and K hold true and let l be some strictly positive integer. For each n and p that may depend on n, define the $U-$processes*

$$V_n^{(k_1,k_2)}(\gamma; l) = \frac{1}{n(n-1)h} \sum_{1 \leq i \neq j \leq n} U_i^{k_1} U_j^{k_2} K_h^l\left(\langle X_i - X_j, \gamma\rangle\right), \quad \gamma \in B_p, \;\; k_1, k_2 \in \{0, 2\}.$$

Then

$$\sup_{\gamma \in B_p} |V_n^{(0,0)}(\gamma; l)| \asymp_{\mathbb{P}} 1, \quad \sup_{\gamma \in B_p} \{1/|V_n^{(2,2)}(\gamma; l)|\} = O_{\mathbb{P}}(1) \;\; \text{and} \;\; \sup_{\gamma \in B_p} |V_n^{(2,0)}(\gamma; l)| = O_{\mathbb{P}}(1).$$

Proof. To simplify the writings, we write $V_n^{(0)}$ (resp. $V_n^{(2)}$) instead of $V_n^{(0,0)}$ (resp. $V_n^{(2,2)}$). First consider the case $k_1 = k_2 = 0$. Hoeffding's decomposition allows us to decompose the centered $U-$processes $hV_n^{(0)}(\gamma; l) - \mathbb{E}[hV_n^{(0)}(\gamma; l)]$ as a sum of two degenerate $U-$processes $V_{1n}^{(0)}(\gamma; l)$ and $V_{2n}^{(0)}(\gamma; l)$, $\gamma \in B_p$, of respective orders 1 and 2 that are indexed by families of functions obtained by finite sums of sets like $\mathcal{F}_{1p}$ and $\mathcal{F}_{2p}$ in Lemma 6.1 above. By Lemma 16 of Nolan and Pollard (1987), deduce that the families indexing $V_{1n}^{(0)}(\gamma;l)$ and $V_{2n}^{(0)}(\gamma;l)$ are families with covering numbers bounded by polynomials in $1/\varepsilon$ with coefficient and order depending on c_1, c_2 and c_3 but independent of n and p. (When $l>1$, K should be replaced by K^l in the definitions of $\mathcal{F}_{1p}$ and $\mathcal{F}_{2p}$, but given the properties of $K(\cdot)$ this has no impact on the conclusion.) Next, by Theorem 2 of Major (2006), $\sup_{\gamma\in B_p}|V_{2n}^{(0)}(\gamma;l)| = O_{\mathbb{P}}(n^{-1}h^{1/2}p^{3/2}\ln n)$; see the proof of our Lemma 3.1 for an example of application of the result of Major (2006). On the other hand, by Theorem 2.14.1 or Theorem 2.14.9 of van der Vaart and Wellner (1996), we have $\sup_{\gamma\in B_p}|V_{1n}^{(0)}(\gamma;l)| = O_{\mathbb{P}}(n^{-1/2}p^{1/2})$. Gathering the rates and using Assumption K-(b,c) we deduce that $V_n^{(0)}(\gamma;l)-\mathbb{E}[V_n^{(0)}(\gamma;l)] = o_{\mathbb{P}}(1)$, uniformly in $\gamma\in B_p$. Now, it remains to show that there exist constants $c_1,c_2>0$ such that $c_1 \leq \mathbb{E}[V_n^{(0)}(\gamma;l)] = \mathbb{E}[h^{-1}K_h^l(\langle X_1-X_2,\gamma\rangle)] \leq c_2$, $\forall\gamma\in B_p$ and h sufficiently small. Using the properties of the Fourier and inverse Fourier transforms, Fubini theorem, the independence of X_1 and X_2 and Plancherel theorem

$$\begin{aligned}
\mathbb{E}[h^{-1}K_h^l(\langle X_1-X_2,\gamma\rangle)] &= (2\pi)^{-1/2}\mathbb{E}\int_{\mathbb{R}}\exp\{it\langle X_1,\gamma\rangle\}\exp\{-it\langle X_2,\gamma\rangle\}\mathcal{F}[K^l](t)dt\\
&= (2\pi)^{1/2}\int_{\mathbb{R}}|\mathcal{F}[f_\gamma](t)|^2\mathcal{F}[K^l](ht)dt\\
&\leq (2\pi)^{1/2}\int_{\mathbb{R}}|\mathcal{F}[f_\gamma](t)|^2dt = (2\pi)^{1/2}\int_{\mathbb{R}}f_\gamma^2(x)dx. \qquad (3.6.4)
\end{aligned}$$

Assumption D-(c)(i) guarantees that $\mathbb{E}[h^{-1}K_h^l(\langle X_1-X_2,\gamma\rangle)]$ is uniformly bounded from above. On the other hand, using the positiveness of $\mathcal{F}[K]$ (hence of $\mathcal{F}[K^l]$), the fact that $\mathcal{F}[K^l]$ is necessarily bounded away from zero on compact intervals, the previous display and Assumption D-(c)(ii), deduce that there exists constants c_3 and c_4 such that $\forall p\geq 1$, $\forall\gamma\in B_p$ and $\forall h\leq 1$ (say),

$$\mathbb{E}[h^{-1}K_h^l(\langle X_1-X_2,\gamma\rangle)] \geq c_3\int_{|t|\leq\epsilon}|\mathcal{F}[f_\gamma](t)|^2dt \geq c_4 > 0.$$

In the case $k_1 = k_2 = 2$, by Assumption D-(b), $\mathbb{E}(V_n^{(2)}(\gamma;l)) \geq \underline{\sigma}^4\mathbb{E}[h^{-1}K_h^l(\langle X_1-X_2,\gamma\rangle)]$, $\forall\gamma$. Next use again Hoeffding's decomposition for $V_n^{(2)}(\gamma;l)-\mathbb{E}(V_n^{(2)}(\gamma;l))$. The degenerate $U-$statistics of order 1 and 2 can be treated with the same arguments as above. Deduce that $1/V_n^{(2)}(\gamma;l)$ is uniformly bounded in probability. The case $k_1=0$ and $k_2=2$ could be handled with similar arguments. ∎

Lemma 6.3 *Suppose Assumptions D and K hold true and let*

$$U_{1n}(\gamma) = \frac{1}{n(n-1)h} \sum_{1 \leq i \neq j \leq n} U_i K_{h,ij}(\gamma), \qquad U_{2n}(\gamma) = \frac{1}{n(n-1)h} \sum_{1 \leq i \neq j \leq n} U_i X_j K_{h,ij}(\gamma).$$

Then

$$\sup_{\gamma \in B_p} |U_{1n}(\gamma)| = O_{\mathbb{P}}(h^{-1/4} n^{-1/2} p^{1/2} \ln^{1/2} n), \quad \sup_{\gamma \in B_p} \|U_{2n}(\gamma)\|_{L^2} = O_{\mathbb{P}}(h^{-1/4} n^{-1/2} p^{1/2} \ln^{3/2} n).$$

Proof. We only prove the statement for $U_{2n}(\gamma)$ from which the uniform rate of $U_{1n}(\gamma)$ could be derived as a particular case. Define

$$\begin{aligned}
\sqrt{h} U_{21n}(\gamma) &= \frac{1}{n(n-1)\sqrt{h}} \sum_{1 \leq i \neq j \leq n} \left[U_i \mathbb{I}_{\{|U_i| \leq H\}} - \mathbb{E}(U_i \mathbb{I}_{\{|U_i| \leq H\}} \mid X_i) \right] X_j K_{h,ij}(\gamma) \\
&= \frac{1}{n} \sum_{1 \leq i \leq n} \left[U_i \mathbb{I}_{\{|U_i| \leq H\}} - \mathbb{E}(U_i \mathbb{I}_{\{|U_i| \leq H\}} \mid X_i) \right] \left\{ \frac{1}{(n-1)\sqrt{h}} \sum_{1 \leq j \leq n} X_j K_{h,ij}(\gamma) \right\} \\
&\quad - \frac{K(0)}{(n-1)\sqrt{h}} \frac{1}{n} \sum_{1 \leq i \leq n} \left[U_i \mathbb{I}_{\{|U_i| \leq H\}} - \mathbb{E}(U_i \mathbb{I}_{\{|U_i| \leq H\}} \mid X_i) \right] X_i \\
&= \frac{1}{n} \sum_{1 \leq i \leq n} \left[U_i \mathbb{I}_{\{|U_i| \leq H\}} - \mathbb{E}(U_i \mathbb{I}_{\{|U_i| \leq H\}} \mid X_i) \right] k_{ni}(\gamma) + o_{\mathbb{P}}(n^{-1/2} h^{1/2}) \\
&= U_{211n}(\gamma) + o_{\mathbb{P}}(n^{-1/2} h^{1/2}),
\end{aligned}$$

with

$$k_{ni}(\gamma) = \frac{1}{(n-1)\sqrt{h}} \sum_{1 \leq j \leq n} X_j K_{h,ij}(\gamma),$$

and let $U_{22n}(\gamma) = U_{2n}(\gamma) - U_{21n}(\gamma)$. If we take $M = n^{1/4} / \ln n$, $U_{22n}(\gamma)$ could be uniformly bounded by taking absolute values, using the fact that K is bounded, applying Cauchy-Schwarz inequality and Markov inequality. The uniform rate $O_{\mathbb{P}}(n^{-1/2})$ follows, see also the proof of Lemma 3.1 for similar arguments. To bound $U_{21n}(\gamma)$ uniformly, we consider a grid of sufficiently close points. The grid can be built with at most n^{3p} points such that any point in B_p is at most at distance $n^{-5/2}$ from a point on the grid. The points $\gamma \in B_p$ outside the grid are handled using the Lipschitz property of the kernel K, see the proofs of Lemmas **??** and 3.1 for similar arguments and more details. For the points γ on the grid we use a Bernstein inequality for Hilbert space independent variables for each point on the grid. The inequality will be applied to $U_{211n}(\gamma)$.

Let us state this exponential inequality such it could be derived for instance from Yurinski (1995), Theorems 3.3.2 or 3.3.4 (see also Bosq (2000), Theorem 2.5) :

if $\xi_1, \cdots, \xi_n$ are independent centered random variables with values in a Hilbert space such that for all natural numbers $l \geq 2$ and some constants $H, B > 0$,

$$\sum_{i=1}^{n} \mathbb{E}\|\xi_i\|^l \leq \frac{B^2}{2}\, l!\; H^{l-2};$$

(H and B may change with n.) Then, for all $s > 0$,

$$\mathbb{P}\left(\left\|\frac{1}{n}\sum_{i=1}^{n}\xi_i\right\| > sB\right) \leq 2\exp\left(-\frac{n^2 s^2}{2[1 + Hns/B]}\right). \tag{3.6.5}$$

We apply this exponential inequality conditionally on $X_1, \cdots, X_n$ for the variables

$$\xi_i = \left[U_i \mathbb{I}_{\{|U_i| \leq M\}} - \mathbb{E}(U_i \mathbb{I}_{\{|U_i| \leq M\}} \mid X_i)\right]\right] k_{ni}(\gamma)$$

for $M = n^{1/4}/\ln n$ and a fixed γ on the grid in B_p, and with the L^2 norm. Note that

$$\mathbb{E}_n[\|\xi_i\|_{L^2}^2] \leq 2\overline{\sigma}^2 \max_{1\leq j\leq n} \|X_j\|_{L^2}^2 \left[\frac{1}{n\sqrt{h}} \sum_{1\leq j\leq n} K_{h,ij}(\gamma)\right]^2,$$

where $\mathbb{E}_n[\cdot] = \mathbb{E}[\cdot \mid X_1, \cdots, X_n]$. Consider the event $\mathcal{E}_n = \{\max_{1\leq j\leq n} \|X_j\|_{L^2} > \ln n\}$ and

$$\mathcal{E}_n(c) = \{\max_{1\leq i\leq n} \sum_{1\leq j\leq n} K_{h,ij}(\gamma) > cn\sqrt{h}\}.$$

By the moment condition imposed on X, $\mathbb{P}(\mathcal{E}_n) \to 0$. On the other hand, for c fixed sufficiently large, $\mathbb{P}(\mathcal{E}_n(c)) \to 0$; see Lemma **??**-(1a,3a). On $\mathcal{E}_n^c \cap \mathcal{E}_n(c)^c$ we could take

$$B^2(\gamma) = \frac{4\overline{\sigma}^2 \ln^2 n}{n^2 h} \sum_{1\leq i,j,k\leq n} K_{h,ij}(\gamma) K_{h,ik}(\gamma).$$

(Without loss of generality, here we consider $K \geq 0$, otherwise it suffice to replace K by $|K|$.) Indeed we have

$$\sum_{i=1}^{n} \mathbb{E}_n[\|\xi_i\|_{L^2}^l] \leq (c\ln n)^{l-2}(n^{1/4}/\ln n)^{l-2} \sum_{i=1}^{n} \mathbb{E}_n[\|\xi_i\|_{L^2}^2] \leq \frac{B^2(\gamma)}{2}\, l!\; H^{l-2},$$

with $H = cn^{1/4}$. By Lemma 6.4,

$$\sup_{\gamma\in B_p} [B^2(\gamma) h^{-1} n^{-1} \ln^{-2} n] \leq C\{1 + o_{\mathbb{P}}(1)\} \qquad \text{for some } 0 < C < \infty,$$

and thus $\mathbb{P}(\{\sup_{\gamma\in B_p} B(\gamma) \geq 2Ch^{1/2} n^{1/2} \ln n\}) \to 0$. Apply inequality (3.6.5) conditionally on $X_1, \cdots, X_n$, on the set $\{\sup_\gamma B(\gamma) \leq 2Ch^{1/2} n^{1/2} \ln n\}$, with

$B = 2Ch^{1/2}n^{1/2}\ln n$, for any point on the grid of B_p with $s = n^{-1}h^{-1/4}[p\ln n]^{1/2}$. Deduce an exponential bound for $U_{211n}(\gamma)$. To obtain an uniform exponential bound for all the points in the grid (there are at most n^{3p}, apply Boole inequality. The uniform bound is simply the bound obtained from (3.6.5) for a fixed γ multiplied by n^{3p}. Deduce the uniform rate $p\ln n \times O_{\mathbb{P}}(h^{-1/4}n^{-1/2}\ln^{1/2} n)$ for $U_{211n}(\gamma)$ conditionally on $X_1, \ldots, X_n$. The uniform rate of $U_{211n}(\gamma)$ is obtained by taking expectation with respect to $X_1, \cdots, X_n$. The uniform rate of $U_{21n}(\gamma)$ follows and the proof is complete. ■

Lemma 6.4 *Under Assumptions D and K,*

$$\sup_{\gamma\in B_p} \frac{1}{n^3h^2} \sum_{1\leq i,j,k\leq n} |K_{h,ij}(\gamma)K_{h,ik}(\gamma)| = C\{1 + o_{\mathbb{P}}(1)\},$$

for some constant $C < \infty$.

Proof. Without loss of generality, consider $K \geq 0$. By elementary calculations,

$$\sup_{\gamma\in B_p} \frac{1}{n^3h^2} \sum_{1\leq i,j,k\leq n} K_{h,ij}(\gamma)K_{h,ik}(\gamma) = \sup_{\gamma\in B_p} \frac{(n-3)!}{n!h^2} \sum_{i\neq j\neq k\neq i} K_{h,ij}(\gamma)K_{h,ik}(\gamma) + O_{\mathbb{P}}(n^{-1}h^{-1}).$$

Hence it suffices to bound in probability the third order $U-$statistics given by

$$g(X_i, X_j, X_k; \gamma) = q(X_i, X_j, X_k; \gamma) - \mathbb{E}[q(X_i, X_j, X_k; \gamma)],$$

where $q(X_i, X_j, X_k; \gamma) = [K_{h,ij}(\gamma)K_{h,ik}(\gamma) + K_{h,ji}(\gamma)K_{h,jk}(\gamma) + K_{h,ki}(\gamma)K_{h,kj}(\gamma)]/3$. For this purpose we apply a Bernstein inequality for $U-$statistics, like the one in Proposition 2.3-(c) of Arcones and Giné (1993). For the sake of completeness, let us recall this inequality. Let $f(V_1, \cdots, V_m)$ symmetric $\|f\|_\infty \leq c < \infty$, $\mathbb{E}[f(V_1, \cdots, V_m) \mid V_1, \cdots, V_{m-1}] = 0$, $\sigma^2 = \mathbb{E}[f^2(V_1, \cdots, V_m)]$. For any $s > 0$,

$$\mathbb{P}\left(\left|\frac{1}{n^{m/2}} \sum_{(i_1,\cdots,i_m)\in I_m^n} f(V_{i_1}, \cdots, V_{i_m})\right| \geq s\right) \leq c_1 \exp\left(-\frac{c_2 s^{2/m}}{\sigma^{2/m} + (cs^{1/m}n^{-1/2})^{2/(m+1)}}\right) \tag{3.6.6}$$

where $I_m^n = \{(i_1, \cdots, i_m) : i_j \in \mathbb{N}, 1 \leq i_j \leq n;\ i_j \neq i_k \text{ if } j \neq k\}$ and c_1, c_2 are constants depending only on m. Next let

$$\pi_{k,m}f(v_1, \cdots, v_k) = (\delta_{v_1} - P)\cdots(\delta_{v_k} - P)P^{m-k}f$$

where δ_v is the Dirac measure and P is the law of V. Let $U_{n,k}(\pi_{k,m}f)$ denote the degenerate $U-$statistic of order k in the Hoeffding decomposition of a $U-$statistics $U_n(f)$ defined by an $m-$variate function f.

Consider the Hoeffding decomposition of the third order $U-$statistics defined by $g(X_i, X_j, X_k; \gamma)$ and let $U_{n,k}(\gamma) = U_{n,k}(\pi_{k,3}g)$ for $k = 3, 2$ and 1. First apply inequality (3.6.6) for $U_{n,k}(\gamma)$. In this case, since $Var(\mathbb{E}[\cdot \mid \cdot]) \leq Var(\cdot)$,

$$\sigma^2 = Var(\pi_{3,3}g) \leq C\mathbb{E}[K^2_{h,ij}(\gamma)K^2_{h,ik}(\gamma)] \leq C'h^2$$

for some constants C, C' independent of γ ; see also equation (3.6.7). Take C'' some large constant independent of γ, and $s = Ch^{1/2}[p\ln n]^{3/2}$ and deduce

$$\mathbb{P}\left(\left|h^{-2}U_{n,3}(\pi_{3,m}g)\right| \geq \frac{C''h^{1/2}[p\ln n]^{3/2}}{n^{3/2}h^2}\right) \leq c_1 \exp\left[-(C''/2)p\ln n\right].$$

Applying this exponential bound on a grid of $n^{7p/2}$ points on the hypersphere $\mathcal{S}^p$ such that any point on the hypersphere is closer than n^{-2} to a point on the grid, and using Boole inequality, deduce that $h^{-2}U_{n,3}(\gamma) = o_{\mathbb{P}}(1)$ uniformly in γ. Next consider the second order $U-$statistics $h^{-1/2}U_{n,2}(\gamma)$ and note that $h^{-1/2}\pi_{2,3}g$ is a bounded bivariate function. Since in this case $\sigma^2 \leq C'h$ for some constant C', apply inequality (3.6.6) with $s = Ch^{1/2}p\ln n$ for some large C, use again a suitable grid and Boole inequality to deduce $h^{-2}U_{n,2}(\gamma) = o_{\mathbb{P}}(1)$ uniformly in γ. Finally, consider the first order $U-$statistics $h^{-1}U_{n,1}(\gamma)$ and note that $h^{-1}\pi_{1,m}g$ is an univariate function bounded by a constant independent of γ. To check this it suffices to note that

$$h^{-1}\mathbb{E}[K_{h,ij}(\gamma)K_{h,ik}(\gamma) \mid X_i] = \mathbb{E}[h^{-1/2}K_{h,ij}(\gamma) \mid X_i]\mathbb{E}[h^{-1/2}K_{h,ik}(\gamma) \mid X_i] \leq C$$

and

$$h^{-1}\mathbb{E}[K_{h,ij}(\gamma)K_{h,ik}(\gamma) \mid X_j] = \mathbb{E}[h^{-1/2}K_{h,ij}(\gamma)\mathbb{E}[h^{-1/2}K_{h,ik}(\gamma) \mid X_i] \mid X_j] \leq C$$

for some constant independent of γ. Since $h^{-1}\pi_{1,m}g$ is bounded, its variance is bounded by a constant times the expectation. The expectation of $h^{-1}\pi_{1,m}g$ could be uniformly bounded by a constant times h (see below). Consequently, apply inequality (3.6.6) with $s = C\sqrt{hp\ln n}$ for some large C. By a suitable grid and Boole inequality to deduce $h^{-2}U_{n,1}(\gamma) = o_{\mathbb{P}}(1)$ uniformly in γ. It remains to bound uniformly $\mathbb{E}[h^{-2}K_{h,ij}(\gamma)K_{h,ik}(\gamma)]$. We can write

$$\begin{aligned}\mathbb{E}[h^{-2}K_{h,ij}(\gamma)K_{h,ik}(\gamma)] &= \frac{1}{2\pi}\mathbb{E}\int_{\mathbb{R}}\int_{\mathbb{R}} e^{it(\langle X_i - X_j,\gamma\rangle}e^{is(\langle X_i - X_k,\gamma\rangle}\mathcal{F}[K](ht)\mathcal{F}[K](hs)dsdt \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} \mathcal{F}[f_\gamma](t+s)\mathcal{F}[f_\gamma](-t)\mathcal{F}[f_\gamma](-s)\mathcal{F}[K](ht)\mathcal{F}[K](hs)dsdt. \quad (3.6.7)\end{aligned}$$

Since

$$\int_{\mathbb{R}} \mathcal{F}[f_\gamma](t+s)\mathcal{F}[f_\gamma](-t)dt = (\mathcal{F}[f_\gamma]*\mathcal{F}[f_\gamma])(s) = \sqrt{2\pi}\int_{\mathbb{R}}\mathcal{F}[f_\gamma^2] = \sqrt{2\pi}\int_{\mathbb{R}} f_\gamma^2 \leq C_1 < \infty,$$

$|\mathcal{F}[f_\gamma](-s)|, |\mathcal{F}[K](ht)|, |\mathcal{F}[K](hs)| \leq 1$, deduce that

$$\sup_p \sup_{\gamma \in B_p} \mathbb{E}[h^{-2} K_{h,ij}(\gamma) K_{h,ik}(\gamma)] \leq C_1.$$

The uniform rate in the statement follows and hence the proof is complete. ∎

3.6.3 Testing for no-effect : proofs of the asymptotic results

Let

$$v_n^2(\gamma_0^{(p)}) = \frac{2}{n(n-1)h} \sum_{j \neq i} \sigma^2_{\gamma_0^{(p)}}(\langle X_j, \gamma_0^{(p)} \rangle) \sigma^2_{\gamma_0^{(p)}}(\langle X_j, \gamma_0^{(p)} \rangle) K_h^2 \left(\langle X_i - X_j, \gamma_0^{(p)} \rangle \right). \tag{3.6.8}$$

Lemma 6.5 *Let Assumptions D, K and hypothesis H_0 hold true. Then $\widehat{\tau}_n^2(\gamma_0^{(p)}) = \tau_n^2(\gamma_0^{(p)})\{1 + o_{\mathbb{P}}(1)\} = v_n^2(\gamma_0^{(p)})\{1 + o_{\mathbb{P}}(1)\}$. Moreover, $\widehat{v}_n^2 = \tau_n^2(\gamma_0^{(p)})\{1 + o_{\mathbb{P}}(1)\}$, with $\widehat{v}_n^2$ defined in (3.3.7), provided that condition (3.3.8) holds true.*

Proof. First let us notice that for any n and any V_{1i}, V_{2i}, $1 \leq i \leq n$, a set of i.i.d. random variables with $\mathbb{E}(V_{1i}^2 + V_{2i}^2) < \infty$ and

$$A_n = \frac{1}{n(n-1)h} \sum_{1 \leq i \neq j \leq n} V_{1i} V_{2j} K(h^{-1} \langle X_i - X_j, \gamma_0^{(p)} \rangle),$$

there exists some constant C (independent of n) such that

$$\begin{aligned} Var(A_n) &\leq \frac{C}{n} Var(V_{1i} V_{2j} h^{-1} K(h^{-1} \langle X_i - X_j, \gamma_0^{(p)} \rangle)) \\ &\leq \frac{C}{nh^2} \mathbb{E}[\zeta_1^2(\langle X_i, \gamma_0^{(p)} \rangle) \zeta_2^2(\langle X_j, \gamma_0^{(p)} \rangle) K^2(h^{-1} \langle X_i - X_j, \gamma_0^{(p)} \rangle)] \\ &\leq \frac{C}{nh^2} \mathbb{E}[\zeta_1^2(\langle X_i, \gamma_0^{(p)} \rangle)] \mathbb{E}[\zeta_2^2(\langle X_j, \gamma_0^{(p)} \rangle)] = \frac{C}{nh^2} \mathbb{E}(V_{1i}^2) \mathbb{E}(V_{2i}^2) \end{aligned} \tag{3.6.9}$$

where $\zeta_l^2(\langle X_i, \gamma_0^{(p)} \rangle) = \mathbb{E}(V_{li}^2 \mid \langle X_i, \gamma_0^{(p)} \rangle)$, $l = 1, 2$. Since $nh^2 \to \infty$, we have $Var(A_n) \to 0$.

Now, to check $\widehat{\tau}_n^2(\gamma_0^{(p)}) = \tau_n^2(\gamma_0^{(p)})\{1 + o_{\mathbb{P}}(1)\}$ take $V_{1i} = V_{2i} = U_i^2$. We have $\mathbb{E}[\widehat{\tau}_n^2(\gamma_0^{(p)}) \mid X_1, \cdots, X_n] = \tau_n^2(\gamma_0^{(p)})$ and

$$\mathbb{E}\{\widehat{\tau}_n^2(\gamma_0^{(p)}) - \tau_n^2(\gamma_0^{(p)})\}^2 = \mathbb{E}\{Var[\widehat{\tau}_n^2(\gamma_0^{(p)}) \mid X_1, \cdots, X_n]\} \leq Var(\widehat{\tau}_n^2(\gamma_0^{(p)})) \to 0. \tag{3.6.10}$$

By the fact that $Var(U \mid X^{(p)})$ is bounded and bounded away from zero almost surely, and the fact that for $l=2$ and $l=4$, $\mathbb{E}[h^{-1}K_h^l(\langle X_1-X_2,\gamma\rangle)]$ is bounded and bounded away from zero $\forall p\geq 1$, $\forall\gamma\in B_p$ and $\forall h\leq 1$, deduce that the expectation of $\tau_n^2(\gamma_0^{(p)})$ stays away from zero and infinity and its variance tends to zero. This together with (3.6.10) allow to conclude that $\widehat{\tau}_n^2(\gamma_0^{(p)}) = \tau_n^2(\gamma_0^{(p)})\{1+o_{\mathbb{P}}(1)\}$. To obtain the same conclusion with $\tau_n^2(\gamma_0^{(p)})$ replaced by $v_n^2(\gamma_0^{(p)})$ it suffices to consider above conditional expectations given $\langle X_1,\gamma_0^{(p)}\rangle,\cdots,\langle X_n,\gamma_0^{(p)}\rangle$. The arguments for $\widehat{v}_n^2$ are similar and hence will be omitted. ■

Proof of Lemma 3.1. Let $M>0$ be a real number that depend on n in a way that will be specified later, define $\eta_i^M = U_i\mathbb{I}(|U_i|\leq M) - \mathbb{E}(U_i\mathbb{I}(|U_i|\leq M)\mid X_i^{(p)})$ and consider the degenerate U-process

$$U_n\widetilde{g} = \frac{1}{n(n-1)}\sum_{j\neq i}\eta_i^M\eta_j^M K_h(\langle X_i-X_j,\gamma\rangle) = \frac{1}{n(n-1)}\sum_{j\neq i}\widetilde{g}((\eta_i^M,X_i),(\eta_j^M,X_j);h,\gamma)$$

defined by the functions $\widetilde{g}$ indexed by h and $\gamma\in\mathcal{S}^p$. By Assumption D and K-(a), the arguments used in Lemma 6.1 above for the class $\mathcal{F}_{1p}$, and Lemma 9.9-(vi) of Kosorok (2008), the bounded family $\mathcal{F}_{3p}=\{\widetilde{g}:\gamma\in\mathcal{S}^p, h>0\}$ has a covering number like in (3.6.3). By Theorem 2 of Major (2006) and its corollary, where we assume without loss of generality that $0\leq K(\cdot)\leq 1$,

$$\mathbb{P}\left(\sup_{\gamma\in\mathcal{S}^p}|U_n\widetilde{g}|\geq\frac{th^{1/2}\ln np^{3/2}}{(n-1)}\right)=\mathbb{P}\left(\sup_{\gamma\in\mathcal{S}^p}\left|\frac{1}{n}\sum_{j\neq i}\frac{\eta_i^M}{M}\frac{\eta_j^M}{M}K_h(\langle X_i-X_j,\gamma\rangle)\right|\geq\frac{th^{1/2}p^{3/2}\ln n}{M^2}\right)$$
$$\leq C_1C_2\exp\left\{-C_3\left(\frac{th^{1/2}p^{3/2}\ln n}{M^2\sigma_M}\right)\right\},\qquad\text{for any } t>0\,,$$

$$\text{provided}\qquad n\sigma_M^2\geq\frac{th^{1/2}p^{3/2}\ln n}{M^2\sigma_M}\geq C_4\left[p+\max(\ln C_2/\ln n,0)\right]^{3/2}\ln\frac{2}{\sigma_M} \tag{3.6.11}$$

where $C_1,\dots C_4>0$ are some constants independent on n, h and M and

$$\sigma_M^2=\sup_{\gamma\in\mathcal{S}^p}\mathbb{E}\left[\left(\frac{\eta_i^M}{M}\right)^2\left(\frac{\eta_j^M}{M}\right)^2K_h^2(\langle X_i-X_j,\gamma\rangle)\right].$$

From Assumption D-(b,c) and using the arguments as in the last part of the proof of Lemma 6.2 above, there is a constant $C>0$ independent of n such that $C^{-1}\leq\sigma_M^2M^4/h\leq C$. Take $M^4=nhp^{-3/2}\ln^{-(1+\delta)}n\to\infty$ with $\delta>0$ arbitrarily small. Hence σ_M^2 is of order $n^{-1}p^{3/2}\ln^{1+\delta}n\to 0$ and for any $t>0$

$$n\sigma_M^2\geq\frac{nh}{CM^4}=C^{-1}p^{3/2}\ln^{1+\delta}n\geq\frac{th^{1/2}p^{3/2}\ln n}{M^2\sigma_M} \tag{3.6.12}$$

provided n is large enough. On the other hand, for any constant $C' > 0$

$$\frac{th^{1/2}p^{3/2}\ln n}{M^2\sigma_M} \geq C^{-1/2}tp^{3/2}\ln n \geq C'p^{3/2}\ln n \to \infty \tag{3.6.13}$$

for any sufficiently large t. Since $(\ln n)^{-1}\ln(2/\sigma_M)$ is bounded by a positive constant as n goes to ∞, Equations (3.6.12) and (3.6.13) show that (3.6.11) is satisfied for our M, with n and t large enough. By Theorem 2 of Major (2006), $U_n\widetilde{g} = O_{\mathbb{P}}\left(n^{-1}h^{1/2}p^{3/2}\ln n\right)$.

Now, it remains to study the tails of U_i, that is we have to derive the orders of the remainder terms

$$2R_{1n}+R_{2n} = \frac{2}{n(n-1)}\sum_{j\neq i}\eta_i^M\xi_j K_h\left(\langle X_i - X_j,\gamma\rangle\right)+\frac{1}{n(n-1)}\sum_{j\neq i}\xi_i\xi_j K_h\left(\langle X_i - X_j,\gamma\rangle\right)$$

where $\xi_i = U_i - \eta_i^M = U_i\mathbb{I}\left(|U_i| > M\right) - \mathbb{E}\left[U_i\mathbb{I}\left(|U_i| > M\right) \mid X_i\right]$. Now, $\mathbb{E}\left[\sup_\gamma |R_{1n}|\right] \leq C\mathbb{E}\left(\left|\eta_i^M\right| |\xi_j|\right) \leq 2C\mathbb{E}\left(|U_i|\right)\mathbb{E}\left(|\xi_j|\right) \leq C'\mathbb{E}\left(|\xi_j|\right)$, and thus by Hölder's and Chebyshev's inequalities

$$\mathbb{E}\left(|\xi_i|\right) \leq 2\mathbb{E}\left[|U_i|\,\mathbb{I}\left(|U_i| > M\right)\right] \leq 2\mathbb{E}^{1/m}\left[|U_i|^m\right]\mathbb{P}^{(m-1)/m}\left[|U_i| > M\right] \leq 2\mathbb{E}\left[|U_i|^m\right] M^{1-m}.$$

Now it remains to choose m sufficiently large such that $M^{1-m} = o\left(n^{-1}h^{1/2}p^{3/2}\ln n\right)$. With Assumption K-(b) and our choice of M, $m > 11$ will be sufficient. Also it is clear that $\sup_\gamma |R_{2n}|$ is of smaller order than $\sup_\gamma |R_{1n}|$.

To prove that the inverse of the variance estimate is bounded in probability, in view of Lemma 6.5, it remains to show that $1/\tau_n^2(\gamma)$, $\gamma \in B_p$, is uniformly bounded in probability. For this recall that $\sigma_p^2(X^{(p)}) \geq \underline{\sigma}^2$ and apply Lemma 6.2. Now the proof is complete. ■

Proof of Lemma 3.2. By definition, $nh^{1/2}Q_n(\gamma_0^{(p)})/\widehat{v}_n(\gamma_0^{(p)}) \leq nh^{1/2}Q_n(\widehat{\gamma}_n)/\widehat{v}_n(\widehat{\gamma}_n) - \alpha_n\mathbb{I}(\widehat{\gamma}_n \neq \gamma_0^{(p)}))$. This implies that

$$0 \leq \mathbb{I}(\widehat{\gamma}_n \neq \gamma_0^{(p)}) \leq nh^{1/2}\alpha_n^{-1}\left\{Q_n(\widehat{\gamma}_n)/\widehat{v}_n(\widehat{\gamma}_n) - Q_n(\gamma_0^{(p)})/\widehat{v}_n(\gamma_0^{(p)})\right\}.$$

From Lemmas 3.1, 6.2 and 6.5,

$$\begin{aligned}\left|\frac{Q_n(\widehat{\gamma}_n)}{\widehat{v}_n(\widehat{\gamma}_n)} - \frac{Q_n(\gamma_0^{(p)})}{\widehat{v}_n(\gamma_0^{(p)})}\right| &\leq 2\max\left[\sup_{\gamma\in B_p}\{1/\widehat{\tau}_n^2(\gamma)\},\ 1/\widehat{v}_n^2\right]\sup_{\gamma\in B_p}|Q_n(\gamma)| \\ &= O_{\mathbb{P}}(n^{-1}h^{-1/2}p^{3/2}\ln n).\end{aligned}$$

Then $\alpha_n p^{-3/2}/\ln n \to \infty$ yields $\mathbb{I}(\widehat{\gamma}_n \neq \gamma_0^{(p)}) = o_{\mathbb{P}}(1)$. Thus $\mathbb{P}(\widehat{\gamma}_n \neq \gamma_0^{(p)}) = \mathbb{E}[\mathbb{I}(\widehat{\gamma}_n \neq \gamma_0^{(p)})] \to 0$. ■

Proof of Theorem 3.3. From Lemma 3.2, the probabilities of the events $\{Q_n(\widehat{\gamma}_n) = Q_n(\gamma_0^{(p)})\}$ and $\{\widehat{v}_n^2(\widehat{\gamma}_n) = \widehat{\tau}_n^2(\gamma_0^{(p)})\}$, with $\widehat{v}_n^2(\cdot)$ defined in (3.3.6), both converge to 1. On the other hand, by Lemma 6.5 above $\widehat{\tau}_n^2(\gamma_0^{(p)}) = \tau_n^2(\gamma_0^{(p)})\{1 + o_{\mathbb{P}}(1)\}$. Moreover, $\widehat{v}_n^2 = \tau_n^2(\gamma_0^{(p)})\{1+o_{\mathbb{P}}(1)\}$, with $\widehat{v}_n^2$ defined in (3.3.7), provided that condition (3.3.8) holds true. Hence it suffices to derive the asymptotic distribution of $nh^{1/2}Q_n(\gamma_0^{(p)})/\tau_n(\gamma_0^{(p)})$ under H_0. For this purpose we use Assumption D-(c)(iii) and proceed like in Theorem 3.3 and Lemma 6.2 of Patilea and Lavergne (2008) ; see also the CLT in Lemma 2 of Guerre and Lavergne (2005). Moreover we use our Lemma 6.2 with $k_1 = k_2 = 0$ and $l = 2$. To be exactly in the case of Lavergne and Patilea (2008), first consider $nh^{1/2}Q_n(\gamma_0^{(p)})/v_n(\gamma_0^{(p)})$ with $v_n(\gamma_0^{(p)})$ defined in (3.6.8). The arguments for the asymptotic normality of $nh^{1/2}Q_n(\gamma_0^{(p)})/v_n(\gamma_0^{(p)})$ are identical to those of Lavergne and Patilea and hence will be omitted. Finally, by Lemma 6.5, $v_n^2(\gamma_0^{(p)}) = \tau_n^2(\gamma_0^{(p)})\{1 + o_{\mathbb{P}}(1)\}$ and the stated result follows. ■

Proof of Theorem 3.4. The proof is based on inequality (3.3.9). Since $\mathbb{E}(U^2 \mid X) \geq \underline{\sigma}^2 + r_n^2\delta^2(X)$, $\mathbb{E}(U \mid X) = r_n\delta(X)$, and $Var(U \mid \langle X, \widetilde{\gamma}\rangle) \geq \underline{\sigma}^2 + r_n^2 Var(\delta(X) \mid \langle X, \widetilde{\gamma}\rangle)$, clearly the variance estimate $\widehat{v}_n^2(\widetilde{\gamma})$ stays away from zero. Hence it suffices to look at the behavior of $Q_n(\gamma)$. By Lemma 2.1-(B) there exists p_0 and $\widetilde{\gamma} \in B_{p_0} \subset \mathcal{S}^{p_0}$ (p_0 and $\widetilde{\gamma}$ independent of n) such that $\mathbb{E}[\delta(X) \mid \langle X, \widetilde{\gamma}\rangle] \neq 0$. Since $\max_{\gamma \in B_p} Q_n(\gamma) \geq Q_n(\widetilde{\gamma})$ for any $p \geq p_0$, it suffices to investigate the rate of $Q_n(\widetilde{\gamma})$. We can write

$$\begin{aligned}
Q_n(\widetilde{\gamma}) &= \frac{1}{n(n-1)h}\sum_{i\neq j} U_i^0 U_j^0 K_h(\langle X_i - X_j, \widetilde{\gamma}\rangle) \\
&\quad + \frac{2r_n}{n(n-1)h}\sum_{i\neq j} U_i^0 \delta(X_j) K_h(\langle X_i - X_j, \widetilde{\gamma}\rangle) \\
&\quad + \frac{r_n^2}{n(n-1)h}\sum_{i\neq j} \delta(X_i)\delta(X_j) K_h(\langle X_i - X_j, \widetilde{\gamma}\rangle) \\
&=: Q_{0n}(\widetilde{\gamma}) + 2r_n Q_{1n}(\widetilde{\gamma}) + r_n^2 Q_{2n}(\widetilde{\gamma}).
\end{aligned}$$

Since $\widetilde{\gamma}$ is fixed (and of finite dimension), $Q_{0n}(\widetilde{\gamma}) = O_{\mathbb{P}}(n^{-1}h^{-1/2})$ (cf. proof of Theorem 3.3). The $U-$statistic $Q_{1n}(\widetilde{\gamma})$ can be decomposed in a degenerate $U-$statistic of order 2 with the rate $O_{\mathbb{P}}(h^{-1}n^{-1}) = O_{\mathbb{P}}(n^{-1/2})$ and the sum average of centered variables

$$\frac{1}{n}\sum_{1\leq i\leq n} U_i^0 \mathbb{E}[\delta(X_j)h^{-1}K_h(\langle X_i - X_j, \widetilde{\gamma}\rangle) \mid X_i].$$

Hence it suffice to bound $v_n^2 = \mathbb{E}\{(U_i^0)^2\mathbb{E}^2[\delta(X_j)h^{-1}K_h(\langle X_i - X_j, \widetilde{\gamma}\rangle) \mid X_i]\}$. There are several assumptions on δ and $f_{\widetilde{\gamma}}$ that could be used. Condition (1) implies that the map $x \mapsto \mathbb{E}[\delta(X_j)h^{-1}K_h(\langle x - X_j, \widetilde{\gamma}\rangle)]$ is bounded. This combined with the bounded conditional variance of U_i^0 yield $v_n^2 \leq c$ for some constant $c > 0$. If condition (2) is met, let $V_i = \langle X_i, \widetilde{\gamma} \rangle$, $v = \langle x, \widetilde{\gamma} \rangle$ and $\overline{\delta}(V_j) = \mathbb{E}[\delta(X_j) \mid V_j]$. Then using the inverse Fourier transform device we have

$$\begin{aligned}
\mathbb{E}[\delta(X_j)h^{-1}K_h(V_i - V_j) \mid X_i = x] &= \mathbb{E}[\delta(X_j)h^{-1}K_h(v - V_j)] \\
&= \mathbb{E}\left[\overline{\delta}(V_j) \int \exp\{it(v - V_j)\}\mathcal{F}[K](ht)dt\right] \\
&= \int_{\mathbb{R}} \exp\{itv\}\mathcal{F}[\overline{\delta} f_{\widetilde{\gamma}}](t)\mathcal{F}[K](ht)dt.
\end{aligned}$$

Use the fact that $\mathcal{F}[\overline{\delta} f_{\widetilde{\gamma}}] \in L^1(\mathbb{R})$, the fact that $\mathcal{F}[K](ht) \to \mathcal{F}[K](0) = 1$ as $h \to 0$ and $|\mathcal{F}[K](ht)| \leq 1$, and Lebesgue dominated convergence theorem to deduce that $v \mapsto \mathbb{E}[\delta(X_j)h^{-1}K_h(v - V_j)]$ is bounded by a constant (and converges to $\overline{\delta}(v)f_{\widetilde{\gamma}}(v)$). Hence v_n^2 is bounded by a constant. Deduce that with any of the conditions (1) or (2), $Q_{1n}(\widetilde{\gamma}) = O_{\mathbb{P}}(n^{-1/2})$. Finally, it is easy to show that $Var[Q_{2n}(\widetilde{\gamma})] \to 0$ (see, e.g.,the proof of equation (26) in Lavergne and Patilea (2008)). It remains to study

$$\mathbb{E}[Q_{2n}(\widetilde{\gamma})] = \int_{\mathbb{R}} |\mathcal{F}[\overline{\delta} f_{\widetilde{\gamma}}]|^2(t)\mathcal{F}[K](ht)dt.$$

If condition (1) holds true, $\overline{\delta} f_{\widetilde{\gamma}} \in L^2(\mathbb{R})$ and by Plancherel theorem and Lebesgue dominated convergence theorem, $\mathbb{E}[Q_{2n}(\widetilde{\gamma})] \to \int_{\mathbb{R}} |\overline{\delta} f_{\widetilde{\gamma}}|^2 > 0$. If condition (2) is met, $\mathcal{F}[\overline{\delta} f_{\widetilde{\gamma}}] \in L^2(\mathbb{R})$ and hence $\overline{\delta} f_{\widetilde{\gamma}} \in L^2(\mathbb{R})$ and continue with the same arguments. Deduce that with any of the Conditions (1) or (2), $Q_{2n}(\widetilde{\gamma}) \asymp O_{\mathbb{P}}(1)$. Collecting the rates, we obtain the result. ■

3.6.4 Testing the functional linear model : proofs of the results

To simplify notation, in this section we write $\|\cdot\|$ instead of $\|\cdot\|_{L^2}$.

Proof of Lemma 4.1. By simple calculations, we have $\widehat{U}_i = U_i - \langle \widehat{b} - b, X_i - \overline{X}_n \rangle - \overline{U}_n$. Let $K_{h,ij}(\gamma)$ be a short notation for $K_h(\langle X_i - X_j, \gamma \rangle)$. We have the following decomposition

$$Q_n(\gamma; \widehat{a}, \widehat{b}) = Q_n(\gamma) - 2V_1(\gamma) - 2V_2(\gamma) + V_3(\gamma) + V_4(\gamma) + 2V_5(\gamma)$$

where

$$V_1 = \frac{\overline{U}_n}{n(n-1)h} \sum_{1 \leq i \neq j \leq n} U_i K_{h,ij}(\gamma), \quad V_2 = \left\langle \widehat{b} - b, \frac{1}{n(n-1)h} \sum_{i \neq j} U_i (X_j - \overline{X}_n) K_{h,ij}(\gamma) \right\rangle,$$

$$V_3 = \frac{1}{n(n-1)h} \sum_{i \neq j} \langle \widehat{b} - b, X_i - \overline{X}_n \rangle \langle \widehat{b} - b, X_j - \overline{X}_n \rangle K_{h,ij}(\gamma)$$

$$V_4 = \frac{\overline{U}_n^2}{n(n-1)h} \sum_{i \neq j} K_{h,ij}(\gamma), \quad V_5 = \overline{U}_n \left\langle \widehat{b} - b, \frac{1}{n(n-1)h} \sum_{i \neq j} (X_j - \overline{X}_n) K_{h,ij}(\gamma) \right\rangle.$$

To prove the rate in the first part of (3.4.14) we will show that

$$\sup_{\gamma \in \mathcal{S}^p} n h^{1/2} |V_j| = o_{\mathbb{P}}(1),$$

for $j = 1$ to $j = 5$. First let us notice that by Fubini Theorem, $\mathbb{E}(\|\overline{X}_n - \mathbb{E}(X)\|^2) = n^{-1} \int_0^1 Var(X(t))dt$ and so $\|\overline{X}_n - \mathbb{E}(X)\| = O_{\mathbb{P}}(n^{-1/2})$.

For V_1 use the fact that $\overline{U}_n = O_{\mathbb{P}}(n^{-1/2})$ and apply Lemma 6.3. Thus there exists $a > 0$ and $0 < \epsilon < 2a(1 - 2\zeta)$ such that

$$\sup_{\gamma \in \mathcal{S}^p} n h^{1/2} |V_1| = n h^{1/2} O_{\mathbb{P}}(n^{-1/2}) O_{\mathbb{P}}(n^{-1/2+\epsilon} p^{1/2} h^{-1/2+a}) = o_{\mathbb{P}}(1).$$

To derive the rate of V_2 let us write

$$\begin{aligned} V_2 &= \left\langle \widehat{b} - b, \frac{1}{n(n-1)h} \sum_{i \neq j} U_i X_j K_{h,ij}(\gamma) \right\rangle \\ &\quad - \left\langle \widehat{b} - b, \overline{X}_n \right\rangle \frac{1}{n(n-1)h} \sum_{i \neq j} U_i K_{h,ij}(\gamma) \\ &= V_{21} - V_{22} \end{aligned}$$

By Cauchy-Schwarz inequality, the rate of $\|\overline{X}_n\|$ and Lemma 6.3, $\sup_{\gamma \in \mathcal{S}^p} |V_{22}| = o_{\mathbb{P}}(n^{-1/2} \|\widehat{b} - b\|) = o_{\mathbb{P}}(n^{-2\rho})$. For the rate of V_{21} use Lemma 6.3 to deduce that

$$\sup_{\gamma \in \mathcal{S}^p} n h^{1/2} |V_2| = o_{\mathbb{P}}(1).$$

For V_3 take absolute values and use Cauchy-Schwarz inequality and triangle inequality :

$$|V_3| \leq \frac{\|\widehat{b} - b\|^2}{n(n-1)h} \sum_{i \neq j} \{\|X_i - \mathbb{E}(X_i)\| + \|\overline{X}_n - \mathbb{E}(X)\|\} \{\|X_j - \mathbb{E}(X_j)\| + \|\overline{X}_n - \mathbb{E}(X)\|\} K_{h,ij}(\gamma).$$

Apply Lemma 6.2 three times and deduce that

$$\sup_{\gamma \in \mathcal{S}^p} n h^{1/2} |V_3| = n h^{1/2} O_{\mathbb{P}}(\|\widehat{b} - b\|^2) O_{\mathbb{P}}(1) = o_{\mathbb{P}}(1).$$

For V_4 apply Lemma 6.2 with $k_1 = 0$, $k_2 = 0$ and $l = 1$ and the rate of $\overline{U}_n$ to deduce

$$\sup_{\gamma\in\mathcal{S}^p} nh^{1/2}|V_4| = nh^{1/2}O_{\mathbb{P}}(n^{-1})O_{\mathbb{P}}(1) = o_{\mathbb{P}}(1).$$

Finally, let us write

$$\begin{aligned} V_5 &= \frac{\overline{U}_n}{n(n-1)h}\sum_{i\neq j}\left\langle \widehat{b}-b, X_j - \mathbb{E}(X)\right\rangle K_{h,ij}(\gamma) \\ &\quad -\left\langle \widehat{b}-b, \overline{X}_n - \mathbb{E}(X)\right\rangle \frac{\overline{U}_n}{n(n-1)h}\sum_{i\neq j} K_{h,ij}(\gamma) \\ &=: V_{51}+V_{52}. \end{aligned}$$

By Cauchy-Schwarz inequality and Lemma 6.2 with $k_1 = 0$, $k_2 = 2$, $l = 1$ and U_j^2 replaced by $\|X_j - \mathbb{E}(X_j)\|$,

$$\sup_{\gamma\in\mathcal{S}^p} nh^{1/2}|V_{51}| = nh^{1/2}O_{\mathbb{P}}(n^{-1/2})O_{\mathbb{P}}(\|\widehat{b}-b\|)O_{\mathbb{P}}(1) = n^{1/2}h^{1/2}O_{\mathbb{P}}(\|\widehat{b}-b\|) = o_{\mathbb{P}}(1).$$

Next, similar arguments for the uniform rate of V_{52}. Deduce that

$$\sup_{\gamma\in\mathcal{S}^p} nh^{1/2}|V_5| = o_{\mathbb{P}}(1).$$

The arguments for the rate in the second part of (3.4.14) are similar and hence will be omitted. ■

Proof of Lemma 4.3. Let $\widehat{g}$ (resp. $\widehat{g}^0$) be the random function defined in (3.4.16) that one would obtain under the null (resp. alternative) hypothesis, that is with covariates X_i and responses $a + \langle b, X_i\rangle + U_i^0$ (resp. $a + \langle b, X_i\rangle + \delta(X_i) + U_i^0$). We can write

$$\begin{aligned} \|\widehat{b}^0 - \widehat{b}\|^2 &= \sum_{j=1}^m (\widehat{b}_j^0 - \widehat{b}_j)^2 = \sum_{j=1}^m \widehat{\theta}_j^{-2}|\langle \widehat{g}-\widehat{g}^0, \widehat{\phi}_j\rangle|^2 \le \sum_{j=1}^m \widehat{\theta}_j^{-2}\|\widehat{g}-\widehat{g}^0\|^2\|\widehat{\phi}_j\|^2 \\ &= r_n^2 \int_0^1 \left(\frac{1}{n}\sum_{i=1}^n \delta(X_i)\{X_i(u) - \overline{X}_n(u)\}\right)^2 du \;\sum_{j=1}^m \widehat{\theta}_j^{-2} =: r_n^2\Gamma_n \sum_{j=1}^m \widehat{\theta}_j^{-2}. \end{aligned}$$

We have

$$\begin{aligned}\mathbb{E}\int_0^1\left(\frac{1}{n}\sum_{i=1}^n \delta(X_i)\{X_i(u)-\mathbb{E}X_i(u)\}\right)^2 du &= \int_0^1 \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n \delta(X_i)\{X_i(u)-\mathbb{E}X_i(u)\}\right)^2 du\\ &= \frac{1}{n^2}\int_0^1\sum_{i=1}^n \mathbb{E}[\delta^2(X_i)\{X_i(u)-\mathbb{E}X(u)\}^2]du\\ &= \frac{1}{n}\int_0^1 \mathbb{E}[\delta^2(X_1)\{X_1(u)-\mathbb{E}X_1(u)\}^2]du\\ &\leq \frac{1}{n}\mathbb{E}^{1/2}[\delta^4(X)]\mathbb{E}[\|X-\mathbb{E}X\|^2],\end{aligned}$$

where for the second equality we used the fact that $\mathbb{E}[\delta(X)\{X-\mathbb{E}X\}]=0$. On the other hand, since $\mathbb{E}[\delta(X)]=0$, by the law of large numbers $\overline{\delta(X)}_n = n^{-1}\sum_{i=1}^n \delta(X_i) = o_{\mathbb{P}}(1)$. Recall that $\|\overline{X}_n-\mathbb{E}(X)\| = O_{\mathbb{P}}(n^{-1/2})$. Deduce that

$$\int_0^1\left(\frac{1}{n}\sum_{i=1}^n \delta(X_i)\{\overline{X_n}(u)-\mathbb{E}X(u)\}\right)^2 \mathrm{d}u = \overline{\delta(X)}_n^2\|\overline{X}_n-\mathbb{E}(X)\|^2 = o_{\mathbb{P}}(n^{-1}),$$

and finally that $\Gamma_n = O_{\mathbb{P}}(n^{-1})$. For the last part, use Theorem 1 of Hall and Horowitz (2007) which provides the rate of $\int_0^1\{\widehat{b}^0(u)-b(u)\}^2\mathrm{d}u$. Next, let us recall that Assumption P-(c) implies $\theta_j \geq cj^{-\alpha}$ for some constant c and thus $\sum_{j=1}^m \theta_j^{-2} = O(n^{(2\alpha+1)/(\alpha+2\beta)}) = o(n)$ provided that $m \asymp n^{1/(\alpha+2\beta)}$. Finally, one can deduce from the equations (5.6) to (5.9) of Hall and Horowitz (2007) that $\sum_{j=1}^m(\theta_j^{-2}-\widehat{\theta}_j^{-2}) = o_{\mathbb{P}}(n)$. Now the proof is complete. ■

Proof of Theorem 4.4. Like in the proof of Theorem 3.4, it suffice to show that $Q_n(\widetilde{\gamma}) \asymp_{\mathbb{P}} r_n^2$ for some fixed p sufficiently large and $\widetilde{\gamma}\in B_p$, where

$$Q_n(\widetilde{\gamma}) = \frac{1}{n(n-1)h}\sum_{i\neq j}\widehat{U}_i\widehat{U}_jK_h(\langle X_i-X_j,\widetilde{\gamma}\rangle).$$

Here $\widehat{U}_i$ are defined as in (3.4.19). There are 15 cross-product terms, all of them similar or identical to those analyzed in the proofs of Theorem 3.4 and Lemma 4.1. For the sake of brevity we omit the details. ■

Chapitre 4

Nonparametric testing for no-effect with functional responses and functional covariates

4.1 Introduction

Consider a sample of independent copies $(U_1, X_1), \cdots, (U_n, X_n)$ of (U, X) where U and X are square-integrable random functions defined on the unit interval. The problem we investigate herein is the test of the hypothesis

$$H_0 : \ \mathbb{E}(U|X) = 0 \quad \text{almost surely (a.s.)} \tag{4.1.1}$$

against the nonparametric alternative $\mathbb{P}[\mathbb{E}(U|X) = 0] < 1$. See for instance Parthasarathy (1967) or Ledoux and Talagrand (1991) for the construction of the conditional expectation of a Hilbert-space valued random variable.

Estimation in models with functional responses has been investigated by Cuevas, Febrero and Fraiman (2002). Since this work, there has been several articles dealing with functional responses, include for example Yao, Müller and Wang (2005), Chiou, Müller and Wang (2004), Antoch and al. (2008), Aguilera, Ocana, and Valderrama (2008), Crambes and Mas (2011).

The problem of testing with functional responses seems to be much less explored. The only contributions all deal with the problem of testing in the functional linear model. Kokoszka and al. (2008) are testing if the slope function is null or not. Horváth and Reeder (2011) consider the problem of changes of the slope function. Gabrys, Horváth and Kokoszka (2010) have built a test to decide if the error term is correlated. All this procedures are unable to detect general departures from the null hypothesis.

The test we introduce herein is based on a dimension reduction idea used by Lavergne and Patilea (2008). Our test is able to detect nonparametric alternatives. The variable U could be heteroscedastic and we do not require the conditional variance of U given X to be known. We do no require the law of the covariate X to be given or to be of a certain type, like for instance Gaussian. The test could be implemented quite easily and performs well in simulations.

The papers is organized as follows. In section 4.2, we introduce notations and the dimension reduction lemma. Section 4.3 is devoted to a first test. We use a U-statistic of order two with a kernel smoothing. This statistic allows us to build the test statistic. In the next section, we refine the U-statistic by introducing the distribution function of a projection of the covariate.

4.2 A dimension reduction lemma

Let us introduce some notation. For any $p \geq 1$, let $\mathcal{S}^p = \{\gamma \in \mathbb{R}^p : \|\gamma\| = 1\}$ denote the unit hypersphere in $\mathbb{R}^p$. Let $L^2[0,1]$ be the space of the square-integrable real-valued functions defined on the unit interval $\langle \cdot, \cdot \rangle$ denote the inner product in $L^2[0,1]$, that is for any $W_1, W_2 \in L^2[0,1]$

$$\langle W_1, W_2 \rangle = \int_0^1 W_1(t) W_2(t) dt.$$

Let $\| \cdot \|_{L^2}$ be the associated norm. Hereafter $\mathcal{R} = \{\rho_1, \rho_2, \cdots\}$ be an arbitrarily fixed orthonormal basis of the function space $L^2[0,1]$, that is $\langle \rho_i, \rho_j \rangle = \delta_{ij}$. Then the response and the predictor processes can be expanded into

$$U(t) = \sum_{j=1}^{\infty} u_j \rho_j(t) \quad \text{and} \quad X(t) = \sum_{j=1}^{\infty} x_j \rho_j(t), \tag{4.2.2}$$

where the random coefficients u_j (resp. x_j) are given by $u_j = \langle U, \rho_j \rangle$ (resp. $x_j = \langle X, \rho_j \rangle$). For a fixed positive integer p and any $W \in L^2[0,1]$, $W^{(p)} \in L^2[0,1]$ will be the projection of X on the subspace generated by the first p elements of the basis $\mathcal{R}$, that is

$$W^{(p)}(t) = \sum_{j=1}^{p} w_j \rho_j(t).$$

By abuse we also identify $W^{(p)}$ with the $p-$dimension random vector $(w_1, \cdots, w_p)$. On the other hand, for any integer $p > 1$ and non random vector $\gamma = (\gamma_1, \cdots, \gamma_p) \in$

$\mathbb{R}^p$, we consider by abuse γ an element in $L^2[0,1]$ with $(\gamma_1, \cdots, \gamma_p, 0, 0, \cdots)$ the coefficients of its expansion and hence $\langle W, \gamma \rangle = \langle W^{(p)}, \gamma \rangle = \sum_{i=1}^{p} x_j \gamma_j$. In the following we will also use $\beta = \sum_{j=1}^{\infty} b_j \rho_j(t)$ to denote a non random element of $L^2[0,1]$.

Our approach relies on the following lemma, an extension of Lemma 2.1 of Lavergne and Patilea (2008) and Theorem 1 in Bierens (1990) to Hilbert space-valued responses and conditioning random variables. For any $\gamma \in \mathcal{S}^p$, let F_γ denote the distribution function of the real-valued variable $\langle X, \gamma \rangle$, that is $F_\gamma(t) = \mathbb{P}(\langle X, \gamma \rangle \leq t)$, $\forall t \in \mathbb{R}$.

Lemma 2.1 *Let $U, X \in L^2[0,1]$ be random functions. Assume that $\mathbb{E}\|U\| < \infty$ and $\mathbb{E}(U) = 0$.*

(A) The following statements are equivalent :

1. $\mathbb{E}(U \mid X) = 0$ *a.s.*
2. $\mathbb{E}\left[\langle U, \mathbb{E}\left(U \mid \langle X, \gamma \rangle\right)\rangle\right] = 0$ *a.s.* $\forall p \geq 1, \forall \gamma \in \mathcal{S}^p$.
3. $\mathbb{E}\left[\langle U, \mathbb{E}\left\{U \mid F_\gamma(\langle X, \gamma \rangle)\right\}\rangle\right] = 0$ *a.s.* $\forall p \geq 1, \forall \gamma \in \mathcal{S}^p$.

(B) Suppose in addition that for any positive real number s,

$$\mathbb{E}(\|U\| \exp\{s\|X\|\}) < \infty. \tag{4.2.3}$$

If $\mathbb{P}[\mathbb{E}(U \mid X) = 0] < 1$, then there exists a positive integer $p_0 \geq 1$ such that for any integer $p > p_0$, the set

$$\mathcal{A} = \{\gamma \in \mathcal{S}^p : \mathbb{E}(U \mid \langle X, \gamma \rangle) = 0 \ a.s.\}$$

has Lebesgue measure zero on the unit hypersphere $\mathcal{S}^p$ and is not dense.

Proof.

(A) We have

$$\begin{aligned} \mathbb{E}(U \mid X) = 0 \ &\Leftrightarrow \ \mathbb{E}(\langle U, \rho_j \rangle \mid X) = 0, \ \forall j \geq 1 \\ &\Leftrightarrow \ \mathbb{E}(\langle U, \rho_j \rangle \mid \langle X, \gamma \rangle) = 0, \ \forall j \geq 1, \forall p \geq 1, \forall \gamma \in \mathcal{S}^p \\ &\Leftrightarrow \ \mathbb{E}(U \mid \langle X, \gamma \rangle) = 0, \ \forall p \geq 1, \forall \gamma \in \mathcal{S}^p \\ &\Leftrightarrow \ \mathbb{E}(U \mid F_\gamma(\langle X, \gamma \rangle)) = 0, \ \forall p \geq 1, \forall \gamma \in \mathcal{S}^p \end{aligned}$$

The first equivalence is due to the fact that $\mathcal{R}$ is an orthonormal basis. We apply for the second equivalence Lemma 2.1 of Patilea, Sanchez and Saumard

(2012), since $\langle U, \rho_j \rangle$ is a real-valued random variable and checks the conditions $\mathbb{E}|\langle U, \rho_j \rangle| < \infty$ by Cauchy-Schwarz and $\mathbb{E}[\langle U, \rho_j \rangle] = 0$. Next, let us notice that $\mathbb{E}[\langle U, \mathbb{E}(U \mid \langle X, \gamma \rangle)\rangle] = \mathbb{E}[\|\mathbb{E}(U \mid \langle X, \gamma \rangle)\|^2] = \mathbb{E}[\|\mathbb{E}(U \mid F_\gamma(\langle X, \gamma \rangle))\|^2] = \mathbb{E}[\langle U, \mathbb{E}\{U \mid F_\gamma(\langle X, \gamma \rangle)\}\rangle]$. Hence the result follows.

(B) First note that

$$\mathcal{A} \subset \bigcap_{j \geq 1} \mathcal{A}_j, \quad \text{where} \quad \mathcal{A}_j = \{\gamma \in \mathcal{S}^p : \mathbb{E}(\langle U, \rho_j \rangle \mid \langle X, \gamma \rangle) = 0 \ a.s.\}.$$

Now, if $\mathbb{P}[\mathbb{E}(U \mid X) = 0] < 1$, then there exists $j \geq 1$ such that $\mathbb{P}[\mathbb{E}(\langle U, \rho_j \rangle \mid X) = 0] < 1$. Finally, apply Lemma 2.1 of Patilea, Sanchez and Saumard (2012) to deduce that $\mathcal{A}_j$, and hence $\mathcal{A}$, have Lebesgue measure zero on the unit hypersphere $\mathcal{S}^p$ and is not dense.

■

Point (A) is a cornerstone for proving the behavior of our test under the null and the alternative hypothesis. Point (B) shows that in applications it will not be difficult to find directions γ able to reveal the failure of the null hypothesis (4.1.1). Under the additional assumption (4.2.3) such directions represent almost all the points on the unit hyperspheres $\mathcal{S}^p$, provided p is sufficiently large. The assumption (4.2.3) is not restrictive for testing purposes. Indeed, if X does not satisfy condition (4.2.3), it suffices to transform X into some variable $W \in L^2[0,1]$ such that the $\sigma-$field generated by W is the same as the one generated by X and the variable W satisfies condition (4.2.3). *

The following new formulation of H_0 are direct consequences of Lemma 2.1.

Corollary 2.2 *Consider a $L^2[0,1]-$valued random variable U such that $\mathbb{E}\|U\| < \infty$. For any $p \geq 1$, let $\omega_p(\gamma, t)$, $\gamma \in \mathbb{R}^p$ and $t \in \mathbb{R}$, be a real-valued function such that $\omega_p(\gamma, \langle X, \gamma \rangle) > 0$ for all $\|\gamma\| = 1$. The following statements are equivalent :*

1. *The null hypothesis (4.1.1) holds true.*
2. *for any $p \geq 1$ and any set $B_p \subset \mathcal{S}^p$ with strictly positive Lebesgue measure in on the unit hypersphere $\mathcal{S}^p$,*

$$\max_{\gamma \in B_p} \mathbb{E}[\langle U, \mathbb{E}(U \mid \langle X, \gamma \rangle)\rangle \omega_p(\gamma, \langle X, \gamma \rangle)] = 0. \tag{4.2.4}$$

*. For instance, given $X = \sum_{j \geq 1} x_j \rho_j$, one may build $w_j = a_j \arctan(x_j)$, where a_j are non random such that $\sum_{j \geq 1} a_j^2 < \infty$ and may use the bounded random function $W = \sum_{j \geq 1} w_j \rho_j \in L^2[0,1]$ (bounded means $\|W\|$ is a bounded random variable) instead of X in the conditioning.

3. *for any $p \geq 1$ and any set $B_p \subset \mathcal{S}^p$ with strictly positive Lebesgue measure in on the unit hypersphere $\mathcal{S}^p$,*

$$\max_{\gamma \in B_p} \mathbb{E}\left[\langle U, \mathbb{E}\left\{U | F_\gamma(\langle X, \gamma\rangle)\right\}\rangle\right] = 0. \tag{4.2.5}$$

This corollary 2.2 establish two new formulations of the null hypothesis. Two tests are proposed , one using the new formulation (4.2.4) in a first part and the other using (4.2.5) in chapter 5 of the manuscript.

4.3 Testing the effect of a functional covariate : a first approach

The new formulation of the null hypothesis, see corollary 2.2 point 2, allows us to build a test statistic.

To avoid handling denominators close to zero, we set the weight function $\omega(\gamma, \cdot)$ in Corollary 2.2 equal to the density of $\langle X, \gamma\rangle$, denoted by $f_\gamma(\cdot)$, which is assumed to exist for any γ. For any $\gamma \in \mathbb{R}^p$, let

$$Q(\gamma) = \mathbb{E}\{\langle U, \mathbb{E}\left(U \mid \langle X, \gamma\rangle\right)\rangle f_\gamma(\langle X, \gamma\rangle)\}.$$

For any $p \geq 1$, let $B_p \subset \mathcal{S}^p$ be a set with strictly positive Lebesgue measure in $\mathcal{S}^p$. By Corollary 2.2, the null hypothesis (4.1.1) holds true if and only if

$$\forall p \geq 1, \quad \max_{\gamma \in B_p} Q(\gamma) = 0. \tag{4.3.1}$$

4.3.1 The test statistic

In view of equation (4.3.1), our goal is to estimate $Q(\gamma)$. With at hand a sample of (U, X), define

$$Q_n\left(\gamma\right) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \langle U_i, U_j\rangle \frac{1}{h} K_h\left(\langle X_i - X_j, \gamma\rangle\right), \quad \gamma \in \mathcal{S}^p,$$

where $K_h\left(\cdot\right) = K\left(\cdot/h\right)$, where $K(\cdot)$ is a kernel and h a bandwidth. In the case of finite dimension covariates, the function $\gamma \mapsto Q_n(\gamma)$ is the statistic considered by Lavergne and Patilea (2008), see also Bierens (1990). For fixed p and $\gamma \in \mathcal{S}^p$, it is well-known that $Q_n\left(\gamma\right)$ has asymptotic centered normal distribution with rate $nh^{1/2}$ under H_0; see for instance Guerre and Lavergne (2005). We will show that the asymptotic normal distribution is preserved even when p grows at a suitable rate with the sample size. On the other hand, Lemma 2.1-(B) indicates that if p is large enough, the maximum of $Q(\gamma)$ over γ stay away from zero.

We will choose a direction γ as the least favorable direction for the null hypothesis H_0 obtained from a penalized criterion based on a standardized version of $Q_n(\gamma)$. Lavergne and Patilea (2008) and Bierens (1990) considered this idea using $Q_n(\gamma)$. Here we use a standardized version of $Q_n(\gamma)$. More precisely, fix some $\beta_0 \in L^2[0,1]$ that could be interpreted as an initial *guess* of an unfavorable direction for H_0. Let b_{0j}, $j \geq 1$, be the coefficients in the expansion of β_0 in the basis $\mathcal{R}$. For any given $p \geq 1$ such that $\sum_{j=1}^{p} b_{0j}^2 > 0$, let

$$\gamma_0^{(p)} = (b_{01}, \cdots, b_{0p}) \Big/ \sqrt{\sum_{j=1}^{p} b_{0j}^2} \ .$$

Let $\widehat{v}_n^2(\cdot)$ be as estimate of the variance of $nh^{1/2}Q_n(\cdot)$. Given $B_p \subset \mathcal{S}^p$ with strictly positive Lebesgue measure in $\mathcal{S}^p$ that contains $\gamma_0^{(p)}$, the least favorable direction γ for H_0 is defined as

$$\widehat{\gamma}_n = \arg\max_{\gamma \in B_p} \left[nh^{1/2} Q_n(\gamma)/\widehat{v}_n(\gamma) - \alpha_n \mathbb{I}_{\left\{\gamma \neq \gamma_0^{(p)}\right\}} \right] , \tag{4.3.2}$$

where $\mathbb{I}_A$ is the indicator function of a set A, and α_n, $n \geq 1$ is a sequence of positive real numbers decreasing to zero at an appropriate rate that depends on the rates of h and p and that will be made explicit below. Let us notice that the maximization used to define $\widehat{\gamma}_n \in \mathcal{S}^p$ is a finite dimension optimization problem. The choice of β_0, and thus of $\gamma_0^{(p)}$, is theoretically irrelevant, it does not affect the asymptotic critical values and the consistency results. However, in practice the choice of β_0 could be related to a priori information of the practitioner on a class of alternatives, like for instance the class of functions depending only on $\langle X, \beta_0 \rangle$.

We will prove that with suitable rates of increase for α_n and p and decrease for h, the probability of the event $\{\widehat{\gamma}_n = \gamma_0^{(p)}\}$ tends to 1 under H_0. Hence $Q_n(\widehat{\gamma}_n)/\widehat{v}_n(\widehat{\gamma})$ behaves asymptotically like $Q_n(\gamma_0^{(p)})/\widehat{v}_n(\gamma_0^{(p)})$, even when p grows with the sample size. Therefore the test statistic we consider is

$$T_n = nh^{1/2} \frac{Q_n(\widehat{\gamma}_n)}{\widehat{v}_n(\widehat{\gamma}_n)} \ . \tag{4.3.3}$$

We will show that an asymptotic α-level test is given by $\mathbb{I}\left(T_n \geq z_{1-a}\right)$, where z_a is the $(1-a)$-th quantile of the standard normal distribution.

4.3.2 Estimating the variance

To find the direction $\widehat{\gamma}_n$ and to build the test statistics (4.3.3), we need to estimate in some way the variance of $nh^{1/2}Q_n(\gamma)$. The approach that is expected not to inflate the variance estimate under the alternatives and thus to guarantee better power small finite samples would involve the estimations of the conditional variance of $nh^{1/2}Q_n(\gamma)$ given X_i's which writes

$$\tau_n^2(\gamma) = \frac{2}{n(n-1)h}\sum_{j\neq i}\sigma_p^2(X_i^{(p)}, X_j^{(p)})K_h^2(\langle X_i - X_j, \gamma\rangle) \ , \tag{4.3.4}$$

where $\sigma_p^2(X_1^{(p)}, X_2^{(p)}) = Var[\langle U_1, U_2\rangle \mid X_1^{(p)}, X_2^{(p)}]$. An estimator can be easily obtained by replacing $\sigma_p^2(\cdot)$ with an estimate in the last expression. In theory, a good solution would be to use a nonparametric estimate of the $2p-$variate function $\sigma_p^2(\cdot)$, but this is practically infeasible given that it expected to let p to grow with the sample size. A simple and convenient solution with high-dimension covariates is then

$$\widehat{\tau}_n^2(\gamma) = \frac{2}{n(n-1)h}\sum_{j\neq i}\langle U_i, U_j\rangle^2 K_h^2(\langle X_i - X_j, \gamma\rangle) \, . \tag{4.3.5}$$

We will show that under suitable conditions : $\sup_{\gamma\in B_p\subset\mathcal{S}^p}\{1/\widehat{\tau}_n^2(\gamma)\} = O_{\mathbb{P}}(1)$. This protect from a too big value of the denominator.

4.3.3 Behavior under the null hypothesis

Let us introduce a first set of assumptions. Below $0_p \in \mathbb{R}^p$ denotes the null vector of dimension p. Moreover, $\mathcal{F}[\cdot]$ denotes the Fourier transform, cf. Rudin (1987).

Assumption D

(a) *The random vectors* $(U_1, X_1), \ldots, (U_n, X_n)$ *are independent draws from the random vector* $(U, X) \in L^2[0,1]\times L^2[0,1]$ *that satisfies* $\mathbb{E}\|U\|^m < \infty$ *for some* $m > 6$.

(b) $\exists\ \underline{\sigma}^2$ *and* C *such that :*

  (i) $0 < \underline{\sigma}^2 \leq Var(\langle U_1, U_2\rangle \mid X_1, X_2)$ *almost surely ;*

  (ii) $\mathbb{E}\left[\|U\|^2 \mid X\right] \leq C$.

(c) *The sets* $B_p \subset \mathcal{S}^p$, $p \geq 1$ *appearing in (4.3.2) are such that :*

  (i) *there exist constant* C_1 *(independent of* n *and* p*) such that* $\forall p \geq 1$ *and* $\forall\gamma \in B_p$, *the variable* $\langle X, \gamma\rangle$ *admits a density* $f_\gamma(\cdot)$ *and*

$$C_1^{-1} \leq \int_{\mathbb{R}} f_\gamma^2 \leq C_1;$$

(ii) *there exists* $C_2, \epsilon > 0$ *such that* $\int_{|x|\leq\epsilon} |\mathcal{F}[f_\gamma]|^2(x)dx \geq C_2$, $\forall p \geq 1$, $\forall \gamma \in B_p$ *;*

(iii) *the initial 'guess'* β_0 *satisfies two conditions :* $\exists C_3$ *such that* $\int_{\mathbb{R}} f^4_{\gamma_0^{(p)}} \leq C_3$, $\forall p \geq 1$ *and* $\exists \alpha > 0$ *such that* $f_{\gamma_0^{(p)}}$ *is* α*-Hölder that is* $\exists C_4$ $\forall x, y$ *such that* $|f_{\gamma_0^{(p)}}(x) - f_{\gamma_0^{(p)}}(y)| \leq C_4|x-y|^\alpha$, $\forall p \geq 1$.

(iv) $B_p \times 0_{p'-p} \subset B_{p'}$, $\forall 1 \leq p < p'$.

Assumption K

(a) *The kernel* K *is a continuous density of bounded variation with strictly positive Fourier transform on the real line. Moreover, the kernel verifies the condition* $\int_{\mathbb{R}} tK(t)dt < +\infty$.

(b) $h \to 0$ *and* $(nh^2)^\alpha / \ln n \to \infty$ *for some* $\alpha \in (0,1)$.

(c) $p \geq 1$ *depends on* n *and there exists a constant* $\lambda > 0$ *such that* $p \ln^{-\lambda} n$ *is bounded.*

Let us comment on these assumptions. The point D-(b) implies that $\exists$ $\underline{\sigma}^2$ and $\overline{\sigma}^2$ such that $0 < \underline{\sigma}^2 \leq Var(\langle U_1, U_2\rangle \mid X_1, X_2) \leq \overline{\sigma}^2 < \infty$ almost surely. Indeed, we have by Cauchy-Schwartz inequality, independency and hypotheses :

$$\begin{aligned}
\mathbb{E}\left[\langle U_1, U_2\rangle^2 \mid X_1, X_2\right] &\leq \mathbb{E}\left[\|U_1\|^2\|U_2\|^2 \mid X_1, X_2\right] \\
&\leq \mathbb{E}^2\left[\|U\|^2 \mid X\right] \\
&\leq C^2.
\end{aligned}$$

The point (b) insure that the estimator and the inverse of the estimator of the variance will not explode. The bounded variation of K, in particular this means K is bounded, is a very mild condition that allows to easily bound covering numbers of families of functions indexed by γ. Continuity and bounded variation guarantee that K can be recovered by inverse Fourier transform. The role of technical assumption of positive Fourier, that is satisfied by triangular, normal, logistic, Student, or Laplace densities, will be explained below. In Assumption K-(c), it is also possible to let p to grow with the sample size at a polynomial rate, instead of the logarithmic rate. However, we will see below that, in theory, this could induce a loss of power for our test. There is a trade off between the moment conditions one imposes for U and the range of rates allowed for the bandwidth and the growth rate for p : higher moments will be needed for wider ranges and faster rates for p. For bandwidths and p satisfying Assumption K-(b,c) it suffices to take $m > 6$ in Assumption D-(a) ; see the proof of Lemma 3.1. Finally, let us comment on Assumption D-(c). On one hand, a key issue in the proof of Lemma 3.1 below and some of the subsequent proofs will be to control the rate of $\mathbb{E}[h^{-1}K_h(\langle X_1 - X_2, \gamma\rangle)]$ uniformly in

$\gamma \in B_p$ as p grows and h decreases with the sample size. To reduce technicalities, we choose the convenient solution that consists in trying to bound this quantity by a constant. Using the Fourier transform and Plancherel theorem, this is guaranteed by a condition like $\int_{\mathbb{R}} f_\gamma^2 \leq C_1$, $\forall \gamma \in B_p$. Such sufficient conditions could be easily achieved for instance if the coefficients x_j of the expansion of X are independent. Then it suffices to fix some $k \geq 1$ such that the density of x_k is bounded and some small c independent of p and to take $B_p = \{(\gamma_1, \cdots, \gamma_k, \cdots, \gamma_p) \in \mathcal{S}^p : |\gamma_k| \geq c\}$. This simple idea could be useful in many other cases than the one of independent coefficients x_j. On the other hand, we have to keep the variance estimate in the denominator of the test statistic (4.3.3) away from zero. For this we have to ensure that $\mathbb{E}[h^{-1}K_h^2(\langle X_1 - X_2, \gamma \rangle)]$ is bounded away from zero uniformly in $\gamma \in B_p$ as p grows and h decreases with the sample size. One easy way to ensure this is to use again the Fourier transform properties, the positiveness of $\mathcal{F}[K]$ and to impose the positive uniform lower bound for the integral of square of $\mathcal{F}[f_\gamma]$ in a neighborhood of the origin, which necessarily induces a uniform lower bound for $\int_{\mathbb{R}} f_\gamma^2$. To summarize, the choice of β_0 and B_p will be decided in the applications and will also depend on the law of X and the choice of the basis $\mathcal{R}$. In view of our extensive simulation experiment, we argue that the choice of B_p is not an issue in applications, one can confidently perform the optimization on the whole hypersphere $\mathcal{S}^p$. Finally, the condition $B_p \times 0_{p'-p} \subset B_{p'}$, $\forall p < p'$, is a mild technical condition that combined with Lemma 2.1-(A) greatly simplifies the proof of the consistency of our test.

The first step is the study of the behavior of the process $Q_n(\gamma)$, $\gamma \in B_p$, under H_0 when p is allowed to increase with the sample size.

Lemma 3.1 *Under Assumptions D and K and if H_0 holds true,*

$$\sup_{\gamma \in B_p \subset \mathcal{S}^p} |Q_n(\gamma)| = O_{\mathbb{P}}(n^{-1}h^{-1/2}p^{3/2}\ln n).$$

Moreover, if $\widehat{\tau}_n^2(\gamma)$ is the estimate defined in equation (4.3.5),

$$\sup_{\gamma \in B_p \subset \mathcal{S}^p} \{1/\widehat{\tau}_n^2(\gamma)\} = O_{\mathbb{P}}(1).$$

We now describe the behavior of $\widehat{\gamma}_n$ under H_0. A suitable rate α_n will make $\widehat{\gamma}_n$ to be equal to $\gamma_0^{(p)}$ with high probability. Under the null, α_n has to grow to infinity sufficiently fast to render the probability of the event $\{\widehat{\gamma}_n = \gamma_0^{(p)}\}$ close to 1. We will see below that, for better detection of alternative hypothesis, α_n should grow as slow as possible. Indeed, slower rates for α_n will allow the selection of directions $\hat{\gamma}_n$ that could be better suited than $\gamma_0^{(p)}$ for revealing the departure from the null hypothesis. The rate of p is also involved in the search of a trade-off for the rate of α_n : larger p renders slower the rate of uniform convergence to zero of $Q_n(\gamma)$, $\gamma \in B_p$, and hence requires larger α_n.

Lemma 3.2 *Under Assumptions D, K, for a positive sequence α_n, $n \geq 1$ such that $\alpha_n/\{p^{3/2}\ln n\} \to \infty$,*

$$\mathbb{P}(\widehat{\gamma}_n = \gamma_0^{(p)}) \to 1, \quad \textit{under } H_0.$$

The following result shows that the asymptotic critical values of our test statistic are standard normal.

Theorem 3.3 *Under the conditions of Lemma 3.2 and if the hypothesis H_0 in (4.1.1) holds true, the test statistic T_n converges in law to a standard normal. Consequently, the test given by $\mathbb{I}(T_n \geq z_{1-a})$, with z_a the $(1-a)-$quantile of the standard normal distribution, has asymptotic level a.*

4.3.4 The behavior under the alternatives

First let us give an intuition on the reason why our test is consistent. Consider the alternative hypothesis

$$H_1: \ \mathbb{P}[\mathbb{E}(U \mid X) = 0] < 1.$$

The way the statistic T_n is constructed guarantees the consistency of our test against H_1. Indeed, we can write

$$\begin{aligned} T_n &= \frac{nh^{1/2}Q_n(\widehat{\gamma}_n)}{\widehat{v}_n(\widehat{\gamma}_n)} \\ &= \max_{\gamma\in B_p}\left\{nh^{1/2}Q_n(\gamma)/\widehat{v}_n(\gamma) - \alpha_n\mathbb{I}_{\{\gamma\neq\gamma_0^{(p)}\}}\right\} + \alpha_n\mathbb{I}_{\{\widehat{\gamma}_n\neq\gamma_0^{(p)}\}} \\ &\geq \frac{\max_{\gamma\in B_p} nh^{1/2}Q_n(\gamma)}{\widehat{v}_n(\gamma_0^{(p)})} - \alpha_n \geq \frac{nh^{1/2}Q_n(\gamma)}{\widehat{v}_n(\gamma_0^{(p)})} - \alpha_n, \quad \forall\gamma \in B_p \subset \mathcal{S}, \end{aligned} \tag{4.3.6}$$

with $\widehat{v}_n(\gamma_0^{(p)})$ equal to $\widehat{\tau}_n(\gamma_0^{(p)})$ defined in (4.3.5). Since $Var(\langle U_1, U_2\rangle \mid \langle X_1 - X_2, \gamma_0^{(p)}\rangle) \geq \underline{\sigma}^2$, it is clear that $1/\widehat{v}_n(\gamma_0^{(p)}) = O_{\mathbb{P}}(1)$. On the other hand, from Lemma 2.1, there exists p_0 and $\widetilde{\gamma} \in B_{p_0}$ such that the expectation of $Q_n(\widetilde{\gamma})$ stays away from zero as the sample size grows to infinity and h decrease to zero. On the other hand, for any $p > p_0$ and any n and h, clearly $\max_{\gamma\in B_p} Q_n(\gamma) \geq Q_n(\widetilde{\gamma})$, because $B_{p_0} \times 0_{p-p_0} \subset B_p$. All these facts show why our test is omnibus, that is consistent against nonparametric alternatives, provided that $p \to \infty$.

To formalize the consistency result, let us fix some $L^2[0,1]$-valued function $\delta(X)$ such that $\mathbb{E}[\delta(X)] = 0$ and $0 < \mathbb{E}[\|\delta(X)\|^4] < \infty$, and some sequence of real numbers r_n that could decrease to zero (the case $r_n \equiv 1$ is also included). Consider the sequence of alternatives

$$H_{1n}: \; U(u) = U^0(u) + r_n\delta(X)(u), \quad n \geq 1, \forall u \in [0,1], \quad \text{with } \mathbb{E}(U^0 \mid X) = 0. \tag{4.3.7}$$

We show below that such directional alternatives can be detected as soon as $r_n^2 nh^{1/2}/\alpha_n$ tends to infinity. This is exactly the same condition as in Lavergne and Patilea (2008). However, in the functional data framework, to obtain the convenient standard normal critical values, we need $1/\alpha_n = o(p^{-3/2}\ln^{-1} n)$. Hence, the rate r_n at which the alternatives H_{1n} tend to the null hypothesis should satisfy $r_n^2 nh^{1/2}/\{p^{3/2}\ln n\} \to \infty$.

Theorem 3.4 *Suppose that*

(a) Assumption D holds true with U replaced by U^0 ;

(b) Assumption K is satisfied;

(c) $\alpha_n/\{p^{3/2}\ln n\} \to \infty$ and r_n, $n \geq 1$ is such that $r_n^2 nh^{1/2}/\alpha_n \to \infty$;

(d) $\mathbb{E}[\delta(X)] = 0$ and $0 < \mathbb{E}[\|\delta(X)\|^4] < \infty$.

Then the test based on T_n is consistent against the sequence of alternatives H_{1n} if there exists $p \geq 1$ and $\widetilde{\gamma} \in B_p$ such that $\mathbb{P}([\delta(X) \mid \langle X, \widetilde{\gamma}\rangle] = 0) < 1$ and one of the following conditions is satisfied :

(i) the density $f_{\widetilde{\gamma}}$ is bounded;

(ii) the function $\mathbb{E}[\delta(X) \mid \langle X, \widetilde{\gamma}\rangle = \cdot]f_{\widetilde{\gamma}}(\cdot)$ is bounded;

(iii) the Fourier transform of $\mathbb{E}[\delta(X)(u) \mid \langle X, \widetilde{\gamma}\rangle = x]f_{\widetilde{\gamma}}(x)$ is integrable on $\mathbb{R} \times [0,1]$ as a function of (t,u).

Let us recall that the existence of p and $\widetilde{\gamma} \in B_p$ such that $\mathbb{P}([\delta(X) \mid \langle X, \widetilde{\gamma}\rangle] = 0) < 1$ is guaranteed by Lemma 2.1.

4.4 Appendix

4.4.1 Rates of convergence : technical lemmas

For ν a probability measure on a sample space, $\mathcal{F}$ a class of functions and $\varepsilon > 0$, let $N(\varepsilon, \mathcal{F}, L^2(\nu))$, denote the covering number, that is the minimal number of balls of radius ε in $L^2(\nu)$ needed to cover $\mathcal{F}$. See Van der Vaart and Wellner (1996) or Kosorok (2008) for the definitions. For real random variables, $A_n \asymp_{\mathbb{P}} B_n$ means that there exists a constant $C > 1$ such that $\mathbb{P}(1/C \leq A_n/B_n \leq C)$ goes to 1 when n grows. In the following $C, C_1, c, c_1, \cdots$ represent constants that may change from line to line.

Lemma 4.1 *For any $p \geq 1$, let*

$$\mathcal{F}_{1p} = \{(v_1, v_2) \mapsto K(h^{-1}\langle v_1 - v_2, \gamma\rangle) : v_1, v_2 \in \mathbb{R}^p, \gamma \in \mathcal{S}^p, h > 0\}$$

and

$$\mathcal{F}_{2p} = \{v \mapsto \mathbb{E}[K(h^{-1}\langle X - v, \gamma\rangle) |: v \in \mathbb{R}^p, \gamma \in \mathcal{S}^p, h > 0\}.$$

If Assumption K-(a) holds, there exist constants $c_1, c_2, c_3 > 0$ such that for any $p \geq 1$ and $0 < \varepsilon < 1$ and any ν_1 probability measure on $\mathbb{R}^p \times \mathbb{R}^p$ and ν_2 probability measure on $\mathbb{R}^p$,

$$N(\varepsilon, \mathcal{F}_{jp}, L^2(\nu_j)) \leq c_1(c_2/\varepsilon)^{c_3 p}, \quad j = 1, 2. \tag{4.4.1}$$

Proof. Since K can be written as a difference of two monotone functions, the result for $\mathcal{F}_{1p}$ is an easy consequence of the Theorem 9.3, Lemmas 9.6 and 9.9 of Kosorok (2008) and Lemma 16 of Nolan and Pollard (1987) ; see also their Lemma 22-(ii). For $\mathcal{F}_{2p}$, use the bound for $\mathcal{F}_{1p}$ and Lemma 20 of Nolan and Pollard (1987). ■

Lemma 4.2 *Let Assumptions D and K hold true and let l be some strictly positive integer. For each n and p that may depend on n, define the $U-$processes*

$$V_n^{(k)}(\gamma; l) = \frac{1}{n(n-1)h} \sum_{1 \leq i \neq j \leq n} \langle U_i, U_j\rangle^k K_h^l\left(\langle X_i - X_j, \gamma\rangle\right), \quad \gamma \in B_p, \;\; k \in \{0, 2\}.$$

Then

$$\sup_{\gamma \in B_p} |V_n^{(0)}(\gamma; l)| \asymp_{\mathbb{P}} 1 \;\; \text{and} \;\; \sup_{\gamma \in B_p} \{1/|V_n^{(2)}(\gamma; l)|\} = O_{\mathbb{P}}(1).$$

Proof. First consider the case $k = 0$. Hoeffding's decomposition allows us to decompose the centered $U-$processes $hV_n^{(0)}(\gamma; l) - \mathbb{E}[hV_n^{(0)}(\gamma; l)]$ as a sum of two degenerate $U-$processes of orders $V_{1n}^{(0)}(\gamma; l)$ and $V_{2n}^{(0)}(\gamma; l)$, $\gamma \in B_p$, of respective orders 1 and 2 that are indexed by families of functions obtained by finite sums of sets like $\mathcal{F}_{1p}$ and $\mathcal{F}_{2p}$ in Lemma 4.1 above. By Lemma 16 of Nolan and Pollard (1987), deduce that the families indexing $V_{1n}^{(0)}(\gamma; l)$ and $V_{2n}^{(0)}(\gamma; l)$ are families with covering numbers bounded by polynomials in $1/\varepsilon$ with coefficient and order depending on c_1, c_2 and c_3 but independent of n and p. (When $l > 1$, K should be replaced by K^l in the definitions of $\mathcal{F}_{1p}$ and $\mathcal{F}_{2p}$, but given the properties of $K(\cdot)$ this has no impact on the conclusion.) Next, by Theorem 2 of Major (2006), $\sup_{\gamma\in B_p} |V_{2n}^{(0)}(\gamma;l)| = O_{\mathbb{P}}(n^{-1}h^{1/2}p^{3/2}\ln n)$; see the proof of our Lemma 3.1 for an example of application of the result of Major (2006). On the other hand, by Theorem 2.14.1 or Theorem 2.14.9 of van der Vaart and Wellner (1996), we have $\sup_{\gamma\in B_p} |V_{1n}^{(0)}(\gamma;l)| = O_{\mathbb{P}}(n^{-1/2}p^{1/2})$. Gathering the rates and using Assumption K-(b,c) we deduce that $V_n^{(0)}(\gamma;l) - \mathbb{E}[V_n^{(0)}(\gamma;l)] = o_{\mathbb{P}}(1)$, uniformly in $\gamma \in B_p$. Now, it remains to show that there exist constants $c_1, c_2 > 0$ such that $c_1 \leq \mathbb{E}[V_n^{(0)}(\gamma;l)] = \mathbb{E}[h^{-1}K_h^l(\langle X_1 - X_2, \gamma\rangle)] \leq c_2$, $\forall \gamma \in B_p$ and h sufficiently small. Using the properties of the Fourier and inverse Fourier transforms, Fubini theorem, the independence of X_1 and X_2 and Plancherel theorem

$$\begin{aligned}
\mathbb{E}[h^{-1}K_h^l(\langle X_1 - X_2, \gamma\rangle)] &= (2\pi)^{-1/2}\mathbb{E}\int_{\mathbb{R}} \exp\{it\langle X_1,\gamma\rangle\}\exp\{-it\langle X_2,\gamma\rangle\}\mathcal{F}[K^l](t)dt \\
&= (2\pi)^{1/2}\int_{\mathbb{R}} |\mathcal{F}[f_\gamma](t)|^2\mathcal{F}[K^l](ht)dt \\
&\leq (2\pi)^{1/2}\int_{\mathbb{R}} |\mathcal{F}[f_\gamma](t)|^2 dt = (2\pi)^{1/2}\int_{\mathbb{R}} f_\gamma^2(x)dx. \qquad (4.4.2)
\end{aligned}$$

Assumption D-(c)(i) guarantees that $\mathbb{E}[h^{-1}K_h^l(\langle X_1 - X_2, \gamma\rangle)]$ is uniformly bounded from above. On the other hand, using the positiveness of $\mathcal{F}[K]$ (hence of $\mathcal{F}[K^l]$), the fact that $\mathcal{F}[K^l]$ is necessarily bounded away from zero on compact intervals, the previous display and Assumption D-(c)(ii), deduce that there exists constants c_3 and c_4 such that $\forall p \geq 1$, $\forall \gamma \in B_p$ and $\forall h \leq 1$ (say),

$$\mathbb{E}[h^{-1}K_h^l(\langle X_1 - X_2, \gamma\rangle)] \geq c_3\int_{|t|\leq\epsilon} |\mathcal{F}[f_\gamma](t)|^2 dt \geq c_4 > 0.$$

In the case $k = 2$, by Assumption D-(b), $\mathbb{E}(V_n^{(2)}(\gamma;l)) \geq \underline{\sigma}^4\mathbb{E}[h^{-1}K_h^l(\langle X_1 - X_2, \gamma\rangle)]$ and the variance of $V_n^{(2)}(\gamma;l)$ can be bounded and shown to converges to zero like in equation (4.4.4) below. Deduce that $1/V_n^{(2)}(\gamma;l)$ is uniformly bounded in probability. ■

4.4.2 Testing for no-effect : proofs of the asymptotic results

Let

$$v_n^2(\gamma_0^{(p)}) = \frac{2}{n(n-1)h}\sum_{j\neq i} \sigma^2_{\gamma_0^{(p)}}(\langle X_i, \gamma_0^{(p)}\rangle, \langle X_j, \gamma_0^{(p)}\rangle)K_h^2\left(\langle X_i - X_j, \gamma_0^{(p)}\rangle\right). \quad (4.4.3)$$

Lemma 4.3 *Let Assumptions D, K and hypothesis H_0 hold true. Then* $\widehat{\tau}_n^2(\gamma_0^{(p)}) = \tau_n^2(\gamma_0^{(p)})\{1+o_{\mathbb{P}}(1)\} = v_n^2(\gamma_0^{(p)})\{1+o_{\mathbb{P}}(1)\}$.

Proof. First let us notice that for any n and any V_{1i}, V_{2i}, $1 \le i \le n$, a set of i.i.d. random variables with $\mathbb{E}(\langle V_{1i}, V_{2j}\rangle^2) < \infty$ and

$$A_n = \frac{1}{n(n-1)h} \sum_{1\le i\neq j\le n} \langle V_{1i}, V_{2j}\rangle K(h^{-1}\langle X_i - X_j, \gamma_0^{(p)}\rangle),$$

there exists some constant C (independent of n) such that

$$\begin{aligned} Var(A_n) &\le \frac{C}{n} Var(\langle V_{1i}, V_{2j}\rangle h^{-1} K(h^{-1}\langle X_i - X_j, \gamma_0^{(p)}\rangle)) \\ &\le \frac{C}{nh^2}\mathbb{E}[\zeta^2(\langle X_i, \gamma_0^{(p)}\rangle, \langle X_j, \gamma_0^{(p)}\rangle) K^2(h^{-1}\langle X_i - X_j, \gamma_0^{(p)}\rangle)] \\ &\le \frac{C}{nh^2}\mathbb{E}[\zeta^2(\langle X_i, \gamma_0^{(p)}\rangle, \langle X_j, \gamma_0^{(p)}\rangle)] = \frac{C}{nh^2}\mathbb{E}(\langle V_{1i}, V_{2i}\rangle^2) \end{aligned} \tag{4.4.4}$$

where $\zeta^2(\langle X_i, \gamma_0^{(p)}\rangle, \langle X_j, \gamma_0^{(p)}\rangle) = \mathbb{E}(\langle V_{1i}, V_{2j}\rangle^2 \mid \langle X_i, \gamma_0^{(p)}\rangle, \langle X_j, \gamma_0^{(p)}\rangle)$. Since $nh^2 \to \infty$, we have $Var(A_n) \to 0$.

Now, to check $\widehat{\tau}_n^2(\gamma_0^{(p)}) = \tau_n^2(\gamma_0^{(p)})\{1+o_{\mathbb{P}}(1)\}$ take $V_{1i} = V_{2i} = U_i$. We have $\mathbb{E}[\widehat{\tau}_n^2(\gamma_0^{(p)}) \mid X_1, \cdots, X_n] = \tau_n^2(\gamma_0^{(p)})$ and

$$\mathbb{E}\{\widehat{\tau}_n^2(\gamma_0^{(p)}) - \tau_n^2(\gamma_0^{(p)})\}^2 = \mathbb{E}\{Var[\widehat{\tau}_n^2(\gamma_0^{(p)}) \mid X_1, \cdots, X_n]\} \le Var(\widehat{\tau}_n^2(\gamma_0^{(p)})) \to 0. \tag{4.4.5}$$

By the fact that $Var(\langle U_1, U_2\rangle \mid X_1^{(p)}, X_2^{(p)})$ is bounded and bounded away from zero almost surely, and the fact that for $l=2$ and $l=4$, $\mathbb{E}[h^{-1}K_h^l(\langle X_1 - X_2, \gamma\rangle)]$ is bounded and bounded away from zero $\forall p \ge 1$, $\forall \gamma \in B_p$ and $\forall h \le 1$, deduce that the expectation of $\tau_n^2(\gamma_0^{(p)})$ stays away from zero and infinity and its variance tends to zero. This together with (4.4.5) allow to conclude that $\widehat{\tau}_n^2(\gamma_0^{(p)}) = \tau_n^2(\gamma_0^{(p)})\{1+o_{\mathbb{P}}(1)\}$. To obtain the same conclusion with $\tau_n^2(\gamma_0^{(p)})$ replaced by $v_n^2(\gamma_0^{(p)})$ it suffices to consider above conditional expectations given $\langle X_1, \gamma_0^{(p)}\rangle, \cdots, \langle X_n, \gamma_0^{(p)}\rangle$. ■

Proof of Lemma 3.1. Let $M > 0$ be a real number that depend on n in a way that will be specified later, define $u_{i,j}^M = \langle U_i, U_j\rangle \mathbb{I}(|\langle U_i, U_j\rangle| \le M)$ and

$$\eta_{ij}^M = u_{i,j}^M - \mathbb{E}(u_{i,j}^M \mid X_i^{(p)}) - \mathbb{E}(u_{i,j}^M \mid X_j^{(p)}) + \mathbb{E}(u_{i,j}^M \mid X_i^{(p)}, X_j^{(p)})$$

and consider the degenerate U-process

$$U_n\widetilde{g} = \frac{1}{n(n-1)} \sum_{j\neq i} \eta_{ij}^M K_h(\langle X_i - X_j, \gamma\rangle) = \frac{1}{n(n-1)} \sum_{j\neq i} \widetilde{g}((\eta_{ij}^M, X_i, X_j); h, \gamma)$$

defined by the functions $\widetilde{g}$ indexed by h and $\gamma \in \mathcal{S}^p$. By Assumption D and K-(a), the arguments used in Lemma 4.1 above for the class $\mathcal{F}_{1p}$, and Lemma 9.9-(vi) of Kosorok (2008), the bounded family $\mathcal{F}_{3p} = \{\widetilde{g} : \|\gamma\| \in \mathcal{S}^p, h > 0\}$ has a covering number like in (4.4.1). By Theorem 2 of Major (2006) and its corollary, where we assume without loss of generality that $0 \le K(\cdot) \le 1$,

$$\mathbb{P}\left(\sup_{\gamma\in\mathcal{S}^p} |U_n\widetilde{g}| \ge \frac{th^{1/2}\ln np^{3/2}}{(n-1)}\right) = \mathbb{P}\left(\sup_{\gamma\in\mathcal{S}^p}\left|\frac{1}{n}\sum_{j\neq i}\frac{\eta_{ij}^M}{M}K_h\left(\langle X_i - X_j, \gamma\rangle\right)\right| \ge \frac{th^{1/2}p^{3/2}\ln n}{M}\right)$$
$$\le C_1C_2\exp\left\{-C_3\left(\frac{th^{1/2}p^{3/2}\ln n}{M\sigma_M}\right)\right\}, \qquad \text{for any } t > 0\ ,$$

$$\text{provided} \qquad n\sigma_M^2 \ge \frac{th^{1/2}p^{3/2}\ln n}{M\sigma_M} \ge C_4\left[p + \max\left(\ln C_2/\ln n, 0\right)\right]^{3/2}\ln\frac{2}{\sigma_M} \tag{4.4.6}$$

where $C_1, \dots C_4 > 0$ are some constants independent on n, h and M and

$$\sigma_M^2 = \sup_{\gamma\in\mathcal{S}^p}\mathbb{E}\left[\left(\frac{\eta_{ij}^M}{M}\right)^2 K_h^2\left(\langle X_i - X_j, \gamma\rangle\right)\right].$$

From Assumption D-(b,c) and using the arguments as in the last part of the proof of Lemma 4.2 above, there is a constant $C > 0$ independent of n such that $C^{-1} \le \sigma_M^2 M^2/h \le C$. Take $M^2 = nhp^{-3/2}\ln^{-(1+\delta)} n \to \infty$ with $\delta > 0$ arbitrarily small. Hence σ_M^2 is of order $n^{-1}p^{3/2}\ln^{1+\delta} n \to 0$ and for any $t > 0$

$$n\sigma_M^2 \ge \frac{nh}{CM^2} = C^{-1}p^{3/2}\ln^{1+\delta} n \ge \frac{th^{1/2}p^{3/2}\ln n}{M\sigma_M} \tag{4.4.7}$$

provided n is large enough. On the other hand, for any constant $C' > 0$

$$\frac{th^{1/2}p^{3/2}\ln n}{M\sigma_M} \ge C^{-1/2}tp^{3/2}\ln n \ge C'p^{3/2}\ln n \to \infty \tag{4.4.8}$$

for any sufficiently large t. Since $(\ln n)^{-1}\ln(2/\sigma_M)$ is bounded by a positive constant as n goes to ∞, Equations (4.4.7) and (4.4.8) show that (**??**) is satisfied for our M, with n and t large enough. By Theorem 2 of Major (2005), $U_n\widetilde{g} = O_{\mathbb{P}}\left(n^{-1}h^{1/2}p^{3/2}\ln n\right)$.

Now, it remains to study the tails of U_i, that is we have to derive the orders of the remainder terms

$$R_n = \frac{1}{n(n-1)}\sum_{j\neq i}\xi_{ij}K_h\left(\langle X_i - X_j, \gamma\rangle\right)$$

where

$$\begin{aligned}
\xi_{ij} &= \langle U_i, U_j\rangle - \eta_{ij}^M \\
&= \langle U_i, U_j\rangle \mathbb{I}\left(|\langle U_i, U_j\rangle| > M\right) \\
&+ \mathbb{E}\left[\langle U_i, U_j\rangle \mathbb{I}\left(|\langle U_i, U_j\rangle| > M\right) \mid X_i^{(p)}\right] \\
&+ \mathbb{E}\left[\langle U_i, U_j\rangle \mathbb{I}\left(|\langle U_i, U_j\rangle| > M\right) \mid X_j^{(p)}\right] \\
&- \mathbb{E}\left[\langle U_i, U_j\rangle \mathbb{I}\left(|\langle U_i, U_j\rangle| > M\right) \mid X_i^{(p)}, X_j^{(p)}\right].
\end{aligned}$$

Now, $\mathbb{E}\left[\sup_\gamma |R_n|\right] \leq C\mathbb{E}\left(|\xi_{ij}|\right)$, and thus by Hölder's, Chebyshev's and Cauchy-Schwartz inequalities

$$\begin{aligned}
\mathbb{E}\left(|\xi_{ij}|\right) &\leq 4\mathbb{E}\left[|\langle U_i, U_j\rangle| \, \mathbb{I}\left(|\langle U_i, U_j\rangle| > M\right)\right] \\
&\leq 4\mathbb{E}^{1/m}\left[|\langle U_i, U_j\rangle|^m\right] \mathbb{P}^{(m-1)/m}\left[|\langle U_i, U_j\rangle| > M\right] \\
&\leq 4\mathbb{E}^2\left[\|U_i\|^m\right] M^{1-m}.
\end{aligned}$$

Now it remains to choose m sufficiently large such that $M^{1-m} = o\left(n^{-1}h^{1/2}p^{3/2}\ln n\right)$. With Assumption K-(b) and our choice of M, $m > 6$ will be sufficient.

To prove that the inverse of the variance estimate is bounded in probability, in view of Lemma 4.3, it remains to show that $1/\tau_n^2(\gamma)$, $\gamma \in B_p$, is uniformly bounded in probability. For this recall that $\sigma_p^2(X^{(p)}) \geq \underline{\sigma}^2$ and apply Lemma 4.2. Now the proof is complete. ■

Proof of Lemma 3.2. By definition, $nh^{1/2}Q_n(\gamma_0^{(p)})/\widehat{v}_n(\gamma_0^{(p)}) \leq nh^{1/2}Q_n(\widehat{\gamma}_n)/\widehat{v}_n(\widehat{\gamma}_n) - \alpha_n\mathbb{I}(\widehat{\gamma}_n \neq \gamma_0^{(p)}))$. This implies that

$$0 \leq \mathbb{I}(\widehat{\gamma}_n \neq \gamma_0^{(p)}) \leq nh^{1/2}\alpha_n^{-1}\left\{Q_n(\widehat{\gamma}_n)/\widehat{v}_n(\widehat{\gamma}_n) - Q_n(\gamma_0^{(p)})/\widehat{v}_n(\gamma_0^{(p)})\right\}.$$

From Lemmas 3.1, 4.2 and 4.3,

$$\begin{aligned}
\left|\frac{Q_n(\widehat{\gamma}_n)}{\widehat{v}_n(\widehat{\gamma}_n)} - \frac{Q_n(\gamma_0^{(p)})}{\widehat{v}_n(\gamma_0^{(p)})}\right| &\leq 2\max\left[\sup_{\gamma\in B_p}\{1/\widehat{\tau}_n^2(\gamma)\},\ 1/\widehat{v}_n^2\right]\sup_{\gamma\in B_p}|Q_n(\gamma)| \\
&= O_{\mathbb{P}}(n^{-1}h^{-1/2}p^{3/2}\ln n).
\end{aligned}$$

Then $\alpha_n p^{-3/2}/\ln n \to \infty$ yields $\mathbb{I}(\widehat{\gamma}_n \neq \gamma_0^{(p)}) = o_{\mathbb{P}}(1)$. Thus $\mathbb{P}(\widehat{\gamma}_n \neq \gamma_0^{(p)}) = \mathbb{E}[\mathbb{I}(\widehat{\gamma}_n \neq \gamma_0^{(p)})] \to 0$. ■

Proof of Theorem 3.3. From Lemma 3.2, the probabilities of the events $\{Q_n(\widehat{\gamma}_n) = Q_n(\gamma_0^{(p)})\}$ and $\{\widehat{v}_n^2(\widehat{\tau}_n) = \widehat{\tau}_n^2(\gamma_0^{(p)})\}$ both converge to 1. On the other hand, by Lemma 4.3 above $\widehat{\tau}_n^2(\gamma_0^{(p)}) = \tau_n^2(\gamma_0^{(p)})\{1 + o_{\mathbb{P}}(1)\}$. Hence it suffices to derive the asymptotic distribution of $nh^{1/2}Q_n(\gamma_0^{(p)})/\tau_n(\gamma_0^{(p)})$ under H_0. To complete the asymptotic normality proof, we apply Theorem 1 of Hall (1984). Define

$$G(Z_i, Z_j) = \mathbb{E}[H_n(Z_i, Z)H_n(Z_j, Z) \mid Z_i, Z_j]$$

where $H_n(Z_1, Z_2) = \langle U_1, U_2\rangle n^{-1}h^{-1/2}K_h(\langle X_1, \gamma_0^{(p)}\rangle - \langle X_2, \gamma_0^{(p)}\rangle)$, $Z = (U, X)$ and $Z_i = (U_i, X_i)$. We have to show condition (2.1) of Hall (1984) :

$$\frac{\mathbb{E}[G^2]}{\mathbb{E}^2[H_n^2]} \to 0 \qquad \text{and} \qquad \frac{n^{-1}\mathbb{E}[H_n^4]}{\mathbb{E}^2[H_n^2]} \to 0 \tag{4.4.9}$$

First, we show the second part of condition 4.4.9. We have

$$\mathbb{E}[H_n^4] = \frac{1}{n^4h^2}\mathbb{E}[\langle U_1, U_2\rangle^4 K_h^4(\langle X_1, \gamma_0^{(p)}\rangle - \langle X_2, \gamma_0^{(p)}\rangle)]$$

which can be bound by $\mathbb{E}[H_n^4] \leq c/(n^4h^2)\mathbb{E}^2\|U_1\|^4$, where c is some constant (independent of n). On the other hand, we have

$$\mathbb{E}[H_n^2] = \frac{1}{n^2h}\mathbb{E}[\langle U_1, U_2\rangle^2 K_h^2(\langle X_1, \gamma_0^{(p)}\rangle - \langle X_2, \gamma_0^{(p)}\rangle)]$$

and $\mathbb{E}[\langle U_1, U_2\rangle^2 h^{-1}K_h^2(\langle X_1, \gamma_0^{(p)}\rangle - \langle X_2, \gamma_0^{(p)}\rangle)]$ can be bound from below by $\underline{\sigma}^2 c$ (see the proof of Lemma 4.2). Due to $nh^2 \to +\infty$, we have $n^{-1}\mathbb{E}[H_n^4]/\mathbb{E}^2[H_n^2] \to 0$. For the first part, let us introduce $V_i = \langle X_i, \gamma_0^{(p)}\rangle$.

$$\begin{aligned}
\mathbb{E}G^2 &\leq \frac{C}{n^4h^2}\mathbb{E}\left[\mathbb{E}^2[K_h(\langle X_i, \gamma_0^{(p)}\rangle - \langle X, \gamma_0^{(p)}\rangle)K_h(\langle X_j, \gamma_0^{(p)}\rangle - \langle X, \gamma_0^{(p)}\rangle) \mid V_i, V_j]\right] \\
&= \frac{C}{n^4h^2}\int\left(\int K_h(v_i - v)K_h(v_j - v)f_{\gamma_0^{(p)}}(v)dv\right)^2 f_{\gamma_0^{(p)}}(v_i)f_{\gamma_0^{(p)}}(v_j)dv_idv_j \\
&= \frac{C}{n^4}\int\left(\int K(s + \frac{v_i - v_j}{h})K(s)f_{\gamma_0^{(p)}}(v_j - hs)ds\right)^2 f_{\gamma_0^{(p)}}(v_i)f_{\gamma_0^{(p)}}(v_j)dv_idv_j \\
&= \frac{Ch}{n^4}\int\left(\int K(s + t)K(s)f_{\gamma_0^{(p)}}(v_j - hs)ds\right)^2 f_{\gamma_0^{(p)}}(v_j + ht)f_{\gamma_0^{(p)}}(v_j)dtdv_j,
\end{aligned}$$

where C is some constant (independent of n). Now, define

$$B = \int\left(\int K(s + t)K(s)ds\right)^2 f_{\gamma_0^{(p)}}^4(v_j)dtdv_j.$$

We have that the difference between the two integrals is $o(h^\alpha)$ using the fact $f_{\gamma_0^{(p)}}$ is α-Hölder with $\alpha > 0$ and the condition of integrability of the kernel K, K-(a). So, we have :

$$\mathbb{E}G^2 \leq \frac{Ch}{n^4}(B + o(h^\alpha))$$

Collecting the results, the first part of the condition 4.4.9 is verified.■

Proof of Theorem 3.4. The proof is based on inequality (4.3.6). Since $\mathbb{E}(\langle U_1, U_2\rangle^2 \mid X_1, X_2) \geq \underline{\sigma}^2 + r_n^4\langle \delta(X_1), \delta(X_2)\rangle^2$, $\mathbb{E}(\langle U_1, U_2\rangle \mid X_1, X_2) = r_n^2\langle \delta(X_1), \delta(X_2)\rangle$, and $Var(\langle U_1, U_2\rangle \mid \langle X_1, \gamma_0^{(p)}\rangle, \langle X_2, \gamma_0^{(p)}\rangle) \geq \underline{\sigma}^2 + r_n^4 Var(\langle \delta(X_1), \delta(X_2)\rangle \mid \langle X_1, \gamma_0^{(p)}\rangle, \langle X_2, \gamma_0^{(p)}\rangle)$, clearly the variance estimate $\widehat{v}_n(\gamma_0^{(p)}))$ stays away from zero. Hence it suffices to look at the behavior of $Q_n(\gamma)$. By Lemma 2.1-(B) there exists p_0 and $\widetilde{\gamma} \in B_{p_0} \subset \mathcal{S}^{p_0}$ (p_0 and $\widetilde{\gamma}$ independent of n) such that $\mathbb{E}[\delta(X) \mid \langle X, \widetilde{\gamma}\rangle] \neq 0$. Since $\max_{\gamma\in B_p} Q_n(\gamma) \geq Q_n(\widetilde{\gamma})$ for any $p \geq p_0$, it suffices to investigate the rate of $Q_n(\widetilde{\gamma})$. We can write

$$\begin{aligned} Q_n(\widetilde{\gamma}) &= \frac{1}{n(n-1)h}\sum_{i\neq j}\langle U_i^0, U_j^0\rangle K_h(\langle X_i - X_j, \widetilde{\gamma}\rangle) \\ &\quad + \frac{2r_n}{n(n-1)h}\sum_{i\neq j}\langle U_i^0, \delta(X_j)\rangle K_h(\langle X_i - X_j, \widetilde{\gamma}\rangle) \\ &\quad + \frac{r_n^2}{n(n-1)h}\sum_{i\neq j}\langle \delta(X_i), \delta(X_j)\rangle K_h(\langle X_i - X_j, \widetilde{\gamma}\rangle) \\ &=: Q_{0n}(\widetilde{\gamma}) + 2r_n Q_{1n}(\widetilde{\gamma}) + r_n^2 Q_{2n}(\widetilde{\gamma}). \end{aligned}$$

Since $\widetilde{\gamma}$ is fixed (and of finite dimension), $Q_{0n}(\widetilde{\gamma}) = O_{\mathbb{P}}(n^{-1}h^{-1/2})$ (cf. proof of Theorem 3.3). The $U-$statistic $Q_{1n}(\widetilde{\gamma})$ can be decomposed in a degenerate $U-$statistic of order 2 with the rate $O_{\mathbb{P}}(h^{-1}n^{-1}) = O_{\mathbb{P}}(n^{-1/2})$ and the sum average of centered variables

$$\frac{1}{n}\sum_{1\leq i\leq n}\langle U_i^0, \mathbb{E}[\delta(X_j)h^{-1}K_h(\langle X_i - X_j, \widetilde{\gamma}\rangle) \mid X_i]\rangle.$$

Hence it suffice to bound $v_n^2 = \mathbb{E}\{\langle U_i^0, \mathbb{E}[\delta(X_j)h^{-1}K_h(\langle X_i - X_j, \widetilde{\gamma}\rangle) \mid X_i]\rangle^2\}$. There are several set of assumptions on δ and $f_{\widetilde{\gamma}}$ that could be used. Condition (i) implies that the map $x \mapsto \mathbb{E}[h^{-1}K_h(\langle x - X_j, \widetilde{\gamma}\rangle)]$ is bounded. This combined with condition (b) on U_i^0 and the finite second order moment of $\delta(X_j)$ yield $v_n^2 \leq c$ for some constant $c > 0$. Similar arguments could be combined with the condition (ii) to obtain the boundedness of v_n^2. If condition (iii) is met,

$$\begin{aligned} v_n^2 &= \mathbb{E}\left(\langle U_i^0, \mathbb{E}[\delta(X_j)h^{-1}K_h(\langle X_i - X_j, \widetilde{\gamma}\rangle) \mid X_i]\rangle^2\right) \\ &\leq \mathbb{E}\left[\|U_i^0\|^2 \|\mathbb{E}[\delta(X_j)h^{-1}K_h(\langle X_i - X_j, \widetilde{\gamma}\rangle) \mid X_i]\|^2\right], \end{aligned}$$

moreover, we have $\|\mathbb{E}[\delta(X_j)h^{-1}K_h(\langle X_i - X_j, \widetilde{\gamma}\rangle) \mid X_i]\| \leq C\mathbb{E}(h^{-1}K_h(\langle X_i - X_j, \widetilde{\gamma}\rangle) \mid X_i)$. Finally, as in lemma 4.2, this quantity is bounded. Deduce that with any of the condition (i) to (iii), $Q_{1n}(\widetilde{\gamma}) = O_{\mathbb{P}}(n^{-1/2})$. Finally, it is easy to show that $Var[Q_{2n}(\widetilde{\gamma})] \to 0$ (see, e.g.,the proof of equation (26) in Lavergne and Patilea (2008)). Let $V_i = \langle X_i, \widetilde{\gamma}\rangle$ and $\overline{\delta}(x,u) = \mathbb{E}[\delta(X_j)(u) \mid V_j = x]$. Then using the inverse Fourier transform device we have for any u

$$\begin{aligned}
\mathbb{E}[\delta(X_j)(u)h^{-1}K_h(V_i - V_j) \mid X_i] &= \mathbb{E}\left[\overline{\delta}(V_j,u)\int \exp\{it(V_i - V_j)\}\mathcal{F}[K](ht)dt \mid V_i\right] \\
&= \int_{\mathbb{R}} \exp\{itV_i\}\mathcal{F}[\overline{\delta}(\cdot,u)f_{\widetilde{\gamma}}(\cdot)](t)\mathcal{F}[K](ht)dt,
\end{aligned}$$

From this and repeated applications of Fubini's theorem we get

$$\begin{aligned}
\mathbb{E}[Q_{2n}(\widetilde{\gamma})] &= \mathbb{E}(\langle\delta(X_i),\delta(X_j)\rangle h^{-1}K_h(\langle X_i - X_j,\widetilde{\gamma}\rangle)) \\
&= \mathbb{E}(\langle\delta(X_i),\delta(X_j)h^{-1}K_h(\langle X_i - X_j,\widetilde{\gamma}\rangle)\rangle) \\
&= \mathbb{E}(\langle\delta(X_i),\mathbb{E}[\delta(X_j)h^{-1}K_h(V_i - V_j) \mid X_i]\rangle) \\
&= \int_0^1 \mathbb{E}\left(\delta(X_i)(u)\int_{\mathbb{R}} \exp\{itV_i\}\mathcal{F}[\overline{\delta}(\cdot,u)f_{\widetilde{\gamma}}(\cdot)](t)\mathcal{F}[K](ht)dt\right)du \\
&= \int_{\mathbb{R}} \|\mathcal{F}[\overline{\delta}(\cdot,u)f_{\widetilde{\gamma}}(\cdot)](t)\|^2\mathcal{F}[K](ht)dt.
\end{aligned}$$

If condition (iii) holds true, for any u, $\mathcal{F}[\overline{\delta}(\cdot,u)f_{\widetilde{\gamma}}(\cdot)](\cdot) \in L^2(\mathbb{R})$ as a function of t. Deduce that $\overline{\delta}(\cdot,u)f_{\widetilde{\gamma}}(\cdot) \in L^2(\mathbb{R})$ as a function of x and by Plancherel theorem and Lebesgue dominated convergence theorem, and since the function $\|\mathcal{F}[\overline{\delta}(\cdot,u)f_{\widetilde{\gamma}}(\cdot)](t)\|^2\mathcal{F}[K](ht)$ is dominated by $\|\mathcal{F}[\overline{\delta}(\cdot,u)f_{\widetilde{\gamma}}(\cdot)](t)\|^2$ which is $L^1(\mathbb{R})$,

$$\mathbb{E}[Q_{2n}(\widetilde{\gamma})] \to \int_{\mathbb{R}} \|\mathcal{F}[\overline{\delta}(\cdot,u)f_{\widetilde{\gamma}}(\cdot)](t)\|^2dt = \iint_{\mathbb{R}\times[0,1]} |\overline{\delta}(x,u)f_{\widetilde{\gamma}}(x)|^2dxdu > 0.$$

Deduce that with any of the conditions (i) to (iii), $Q_{2n}(\widetilde{\gamma}) \asymp O_{\mathbb{P}}(1)$. Collecting the rates, we obtain the result. ■

Chapitre 5

Nonparametric testing for no-effect with functional responses and functional covariates

5.1 Introduction

There has been substantial recent work on the methodology of regression analysis with functional data where predictors, responses, or both of them can be viewed as random functions. Functional data arise in many applications, the monograph of Ramsay and Silverman (2005) provides many compelling examples. In this paper we focus on the case where both the response and the predictor (or covariate) are random elements taking values in a space of functions. The functional linear model is the benchmark approach, see Chiou, Müller and Wang (2004), Yao, Müller and Wang (2005), Gabrys, Horváth and Kokoszka (2010) and the references therein. Recently, alternative nonparametric approaches have been considered ; see Ferraty et al. (2011), Lian (2011), Ferraty, Van Keilegom and Vieu (2012).

An important step in the statistical modeling is the goodness-of-fit of the model considered, for instance the functional linear model. To our best knowledge only the papers of Chiou and Müller (2007) and Kokoszka *et al.* (2008) investigate the problem of goodness-of-fit. Chiou and Müller (2007) introduced diagnostics of the functional regression fit using plots of functional principal components (FPC) scores of the response and the covariate. They also used residuals versus fitted values FPC scores plots. (The FPC are the random coefficients in the Karhunen-Loève expansions.) It is easy to understand that such two-dimension plots could not capture all types of effects of the covariate on the response, such for instance the effect of the interactions of the covariate FPC. Kokoszka *et al.* (2008) used the response and covariate FPC scores to build a test statistic with χ^2 distribution under the null hypothesis of no linear effect. Again, by construction, the test of Kokoszka *et al.* cannot detect any nonlinear alternative. When little is known about the structure of the data, it is preferable to allow for flexible, nonparametric, alternatives for the goodness-of-fit test. Moreover, when proceeding to nonparametric estimation of the link between the response and the predictor, one should also check whether the predictor has an effect of the response or not.

Formally, the statistical issue we address in this paper could be formulated as follows. Consider a sample of independent copies $(U_1, X_1), \cdots, (U_n, X_n)$ of (U, X) where U and X takes values in some separable Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$. Without loss of generality we may suppose that U has zero expectation. The problem is to build a statistical test of the hypothesis of no-effect of U on X, that is

$$H_0 : \ \mathbb{E}\,(U|X) = 0 \quad \text{almost surely (a.s.),} \tag{5.1.1}$$

against the nonparametric alternative $\mathbb{P}[\mathbb{E}\,(U|X) = 0] < 1$.[*] Since $\mathcal{H}_1$ or $\mathcal{H}_2$ could be of finite dimension, for instance the real line, this framework covers all the common situations involving functional data. However, our focus of interest will be on the case functional response and functional covariate.

The goodness-of-fit or no-effect against *nonparametric alternatives* has been very little explored in functional data context. In the case of scalar response, Delsol, Ferraty and Vieu (2011) proposed a testing procedure adapted from the approach of Härdle and Mammen (1993). However, their procedure involves smoothing in the functional space and requires quite restrictive conditions which make it difficult to apply to real data situations. Patilea, Sánchez-Sellero and Saumard (2012) and García-Portugués, González-Manteiga and Febrero-Bande (2012) proposed alternative nonparametric goodness-of-fit tests for scalar response and functional covariate using one dimension projections of the covariate. Such projection-based methods are much less restrictive and perform well in applications. To our best knowledge, no nonparametric statistical test of no-effect or goodness-of-fit is available when both the response and the covariate are functional.

Our test is based on the remark that checking the no-effect of the functional covariate is equivalent to checking the nullity of the conditional expectation of the response given a sufficiently rich set of projections of the covariate. Such projections could be on elements of norm 1 from finite-dimension subspaces of the Hilbert space where the covariate takes values. Then, the idea is to search a finite-dimension element of norm 1 that is, in some sense, the least favorable for the null hypothesis. With at hand such a least favorable direction, it remains to check the nullity of the

∗. See for instance Parthasarathy (1967) for the construction of the expectation and conditional expectation of a Hilbert-space valued random variable.

conditional expectation of the functional response given the scalar product between the covariate and the selected direction. Patilea, Sánchez-Sellero and Saumard (2012) used a similar idea with scalar responses. We follow these steps using a nearest neighbors (NN) smoothing approach. As a result, our new test statistic is a quadratic form involving univariate NN smoothing and the asymptotic critical values are given by the standard normal law. When the response is univariate, our statistic is related but different from the one introduced by Patilea, Sánchez-Sellero and Saumard (2012). By construction, the test is able to detect nonparametric alternatives. The responses could be heteroscedastic with conditional variance of unknown form. The law of the covariate does not need to be known.

The paper is organized as follows. In section 5.2 we introduce the main notation and we derive the fundamental lemma of our approach. This lemma states the equivalence between condition (5.1.1) and the nullity of the conditional expectation of U given a sufficiently rich set of projections of X. In section 5.3 we introduce the test statistic for testing the no-effect of X on U when U is observed. We prove that, under mild technical assumptions, the induced test has one-sided standard normal critical values and it is consistent against *any* type of fixed alternatives and against sequences of directional alternatives approaching the null hypothesis at a suitable rate. The allowed rates are almost the same as those obtained in parametric model checks based on smoothing with *univariate* covariate, see for instance Guerre and Lavergne (2005). Clearly, our test procedure could be also applied in the case where the sample of U is not observed and has to be estimated, for instance as the residual of a regression. Under suitable regularity conditions ensuring that the sample values of U are estimated sufficiently accurate, the test statistic will still have standard normal critical values. In section 5.3.4 we address some practical aspects, namely the choice of the orthonormal basis for decomposing the functional covariate and the finite sample approximation of the critical values. As an example we consider the theoretical basis given by the orthonormal functions of the covariance operator of X and we prove that the asymptotic behaviour of our test statistic does not change when this basis is replaced by its empirical counterpart. Moreover, we propose a simple wild bootstrap procedure to approximate the critical values of our test statistic with small samples. In section 5.4 we illustrate our theoretical findings through an extensive simulation experiment and we present a real data application. In particular we compare our test with the one proposed by Kokoszka *et al.* (2008). We conclude that the test could be easily implemented and performs well in applications. The proofs are relegated to the appendix.

5.2 A dimension reduction lemma

In order to simplify the presentation and without loss of generality, hereafter we focus on the case where the Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ are both equal to the space of square-integrable random functions defined on the unit interval.

Let us introduce some notation. For any $p \geq 1$, let $\mathcal{S}^p = \{\gamma \in \mathbb{R}^p : \|\gamma\| = 1\}$ denote the unit hypersphere in $\mathbb{R}^p$. Let $L^2[0,1]$ be the space of the square-integrable real-valued functions defined on the unit interval $\langle \cdot, \cdot \rangle$ denote the inner product in $L^2[0,1]$, that is for any $W_1, W_2 \in L^2[0,1]$

$$\langle W_1, W_2 \rangle = \int_0^1 W_1(t) W_2(t) dt.$$

Let $\|\cdot\|$ be the associated norm. Hereafter, if not stated differently, $\mathcal{R} = \{\rho_1, \rho_2, \cdots\}$ will be an arbitrarily fixed orthonormal basis of the function space $L^2[0,1]$, that is $\langle \rho_i, \rho_j \rangle = \delta_{ij}$. Then the response and the predictor processes can be expanded into

$$U(t) = \sum_{j=1}^{\infty} u_j \rho_j(t) \quad \text{and} \quad X(t) = \sum_{j=1}^{\infty} x_j \rho_j(t), \tag{5.2.2}$$

where the random coefficients u_j (resp. x_j) are given by $u_j = \langle U, \rho_j \rangle$ (resp. $x_j = \langle X, \rho_j \rangle$). For a fixed positive integer p and $X \in L^2[0,1]$, $X^{(p)} \in L^2[0,1]$ will be the projection of X on the subspace generated by the first p elements of the basis $\mathcal{R}$, that is

$$X^{(p)}(t) = \sum_{j=1}^{p} x_j \rho_j(t).$$

By abuse we identify $X^{(p)}$ with the $p-$dimension random vector $(x_1, \cdots, x_p)$. Hence, for any integer $p \geq 1$ and non random vector $\gamma = (\gamma_1, \cdots, \gamma_p) \in \mathbb{R}^p$, we write $\langle X^{(p)}, \gamma \rangle = \sum_{i=1}^p x_j \gamma_j$. Moreover, we identify $\gamma \in \mathbb{R}^p$ with $\sum_{j=1}^p \gamma_j \rho_j(t) \in L^2[0,1]$ and hence we also write $\langle X, \gamma \rangle = \langle X^{(p)}, \gamma \rangle$. In the following we will also use $\beta = \sum_{j=1}^\infty b_j \rho_j(t)$ to denote a non random element of $L^2[0,1]$.

Our approach relies on the following lemma, an extension of Lemma 2.1 of Lavergne and Patilea (2008) and Theorem 1 in Bierens (1990) to Hilbert space-valued responses and conditioning random variables. For any $\gamma \in \mathcal{S}^p \subset \mathbb{R}^p$, let F_γ denote the distribution function (d.f.) of the real-valued variable $\langle X, \gamma \rangle$, that is $F_\gamma(t) = \mathbb{P}(\langle X, \gamma \rangle \leq t)$, $\forall t \in \mathbb{R}$.

Lemma 2.1 *Let $U, X \in L^2[0,1]$ be random functions. Assume that $\mathbb{E}\|U\| < \infty$ and $\mathbb{E}(U) = 0$.*

(A) The following statements are equivalent :

1. $\mathbb{E}(U \mid X) = 0$ *a.s.*
2. $\mathbb{E}\left[\langle U, \mathbb{E}\left(U \mid \langle X, \gamma\rangle\right)\rangle\right] = 0$ *a.s.* $\forall p \geq 1, \forall \gamma \in \mathcal{S}^p$.
3. $\mathbb{E}\left[\langle U, \mathbb{E}\left\{U \mid F_\gamma(\langle X, \gamma\rangle)\right\}\rangle\right] = 0$ *a.s.* $\forall p \geq 1, \forall \gamma \in \mathcal{S}^p$.

(B) Suppose in addition that for any positive real number s,

$$\mathbb{E}(\|U\| \exp\{s\|X\|\}) < \infty. \tag{5.2.3}$$

If $\mathbb{P}[\mathbb{E}(U \mid X) = 0] < 1$, then there exists a positive integer p_0 such that for any integer $p \geq p_0$, the set

$$\mathcal{A} = \{\gamma \in \mathcal{S}^p : \mathbb{E}(U \mid \langle X, \gamma\rangle) = 0 \ a.s.\} = \{\gamma \in \mathcal{S}^p : \mathbb{E}(U \mid F_\gamma(\langle X, \gamma\rangle)) = 0 \ a.s.\}$$

has Lebesgue measure zero on the unit hypersphere $\mathcal{S}^p$ and is not dense.

Point (A) is a cornerstone for proving the behavior of our test under the null and the alternative hypotheses. Point (B) shows that in applications it will not be difficult to find directions γ able to reveal the failure of the null hypothesis (5.1.1) since, under the very mild † conditions, such directions represent almost all the points on the unit hyperspheres $\mathcal{S}^p$, provided p is sufficiently large.

Let

$$Q(\gamma) = \mathbb{E}[\langle U \ , \mathbb{E}\{U \mid F_\gamma(\langle X, \gamma\rangle)\}\rangle] \tag{5.2.4}$$

The following new formulation of H_0 is a direct consequences of Lemma 2.1 above.

Corollary 2.2 *Consider a $L^2[0,1]-$valued random variable U such that $\mathbb{E}\|U\| < \infty$. The following statements are equivalent :*

1. *The null hypothesis (5.1.1) holds true.*
2. *for any $p \geq 1$ and any set $B_p \subset \mathcal{S}^p$ with strictly positive Lebesgue measure on the unit hypersphere $\mathcal{S}^p$,*

$$\forall p \geq 1, \quad \max_{\gamma \in B_p} Q(\gamma) = 0. \tag{5.2.5}$$

5.3 Testing the effect of a functional covariate

We introduce a general approach for nonparametric testing the no-effect of a functional covariate X on a functional random variable U based on the characterization (5.2.5) of the null hypothesis.

†. If X does not satisfy condition (5.2.3), it suffices to transform X into some variable $W \in L^2[0,1]$ such that the $\sigma-$field generated by W is the same as the one generated by X and the variable W satisfies condition (5.2.3).

5.3.1 The test statistic

In view of equation (5.2.5), our goal is to estimate $Q(\gamma)$. With at hand a sample of (U, X), define

$$Q_n(\gamma) = \frac{1}{n(n-1)} \sum_{1 \le i \ne j \le n} \langle U_i, U_j \rangle \frac{1}{h} K_h \left(F_{\gamma,n}(\langle X_i, \gamma \rangle) - F_{\gamma,n}(\langle X_j, \gamma \rangle) \right), \quad \gamma \in \mathcal{S}^p,$$

where $K_h(\cdot) = K(\cdot/h)$, $K(\cdot)$ is a kernel, h the bandwidth, and $F_{\gamma,n}$ is the empirical d.f. of the sample $\langle X_1, \gamma \rangle, \cdots, \langle X_n, \gamma \rangle$. ‡

The statistic $Q_n(\gamma)$ is related to statistics considered by Fan and Li (1996) and Zheng (1996) for checks of parametric regressions for finite dimension data. See also Patilea, Sánchez-Sellero and Saumard (2012) for the extension of this type of statistics to testing the goodness-of-fit of functional linear model. The statistics considered by all these authors are based on a Nadaraya-Watson regression estimator. Here we use the nearest neighbor (NN) approach of Stute (1984) and hence our new statistic is more in the spirit of the one introduced by Stute and González Manteiga (1996) to test simple linear models with scalar outcome and covariate and homoscedastic error term. Herein we allow for heteroscedasticity of unknown form and hence, in the particular case where U and X are scalar, we extend the framework of Stute and González Manteiga (1996).

The idea of using projections of the covariates was also considered by Lavergne and Patilea (2008) ; see also Bierens (1990), Cuesta-Albertos *et al.* (2007), Cuesta-Albertos, Fraiman and Ransford (2007). The extension of the scope to functional responses seems to be new.

Under H_0, by the Central Limit Theorem (CLT) for degenerate $U-$statistics, for fixed p and $\gamma \in \mathcal{S}^p$, $nh^{1/2} Q_n(\gamma)$ has asymptotic centered normal distribution. Here we use the CLT in Theorem 5.1 in de Jong (1987). We will show de Jong CLT still applies and the asymptotic normal distribution is preserved even when p grows at a suitable rate with the sample size. On the other hand, Lemma 2.1-(B) indicates that if p is large enough, the maximum of $Q(\gamma)$ over γ stays away from zero under the alternative hypothesis and this will guarantee consistency against any departure from H_0.

The statistic $Q_n(\gamma)$ is expected to be close to $Q(\gamma)$ uniformly in γ, provided p increases suitably. Then a natural idea would be to build a test statistic using the maximum of $Q_n(\gamma)$ with respect to γ. However, like in the finite dimension

‡. Ties in the values $\langle X_i, \gamma \rangle$, $1 \le i \le n$, could be broken by comparing indices, that is if $\langle X_i, \gamma \rangle = \langle X_j, \gamma \rangle$, then we define $F_{\gamma,n}(\langle X_i, \gamma \rangle) < F_{\gamma,n}(\langle X_j, \gamma \rangle)$ if $i < j$. However, for simplicity in our assumptions below we will assume that the $\langle X_i, \gamma \rangle$'s have continuous distribution for all γ.

covariate case, under H_0 one expects $Q_n(\gamma)$ to converges to zero for any p and γ and thus the objective function of the maximization problem to be flat. Therefore we will choose a direction γ as the least favorable direction for the null hypothesis H_0 obtained from a penalized criterion based on a standardized version of $Q_n(\gamma)$; see also Lavergne and Patilea (2008) and Bierens (1990) for related approaches. More precisely, fix some infinite-dimension vector $(b_{01}, b_{02}, \cdots) \in \mathbb{R}^\infty$ with $\sum_{j=1}^\infty b_{0j}^2 < \infty$. Such a vector could be interpreted as an initial *guess* of an unfavorable direction for H_0. For any given $p \geq 1$ such that $\sum_{j=1}^p b_{0j}^2 > 0$, let

$$\gamma_0^{(p)} = \frac{(b_{01}, \cdots, b_{0p})}{\|(b_{01}, \cdots, b_{0p})\|} \in \mathcal{S}^p ,$$

where here $\|\cdot\|$ denotes the norm in $\mathbb{R}^p$.

Let

$$\widehat{v}_n^2(\gamma) = \frac{2}{n(n-1)h} \sum_{j \neq i} \langle U_i, U_j \rangle^2 K_h^2 \left(F_{\gamma,n}(\langle X_i, \gamma \rangle) - F_{\gamma,n}(\langle X_j, \gamma \rangle) \right), \tag{5.3.1}$$

be an estimate of the variance of $nh^{1/2} Q_n(\gamma)$, $\gamma \in \mathcal{S}^p$. Given $B_p \subset \mathcal{S}^p$ with positive Lebesgue measure in $\mathcal{S}^p$ and that contains $\gamma_0^{(p)}$, the least favorable direction γ for H_0 is defined by

$$\widehat{\gamma}_n = \arg\max_{\gamma \in B_p} \left[nh^{1/2} Q_n(\gamma)/\widehat{v}_n(\gamma) - \alpha_n \mathbb{I}_{\left\{\gamma \neq \gamma_0^{(p)}\right\}} \right] , \tag{5.3.2}$$

where $\mathbb{I}_A$ is the indicator function of a set A, and α_n, $n \geq 1$ is a sequence of positive real numbers decreasing to zero at an appropriate rate that depends on the rates of h and p and that will be made explicit below. Using a standardized version of $Q_n(\gamma)$ avoids scaling α_n according to the variability of the observations. Let us notice that the maximization used to define $\widehat{\gamma}_n \in B_p \subset \mathcal{S}^p$ is a finite dimension optimization problem. The choice of $\gamma_0^{(p)}$ will be shown to be theoretically irrelevant, it will not affect the asymptotic critical values and the consistency results. Practical aspects related to the choice of $\gamma_0^{(p)}$ and B_p will be discussed in section 5.3.4.

We will prove that with suitable rates of increase for α_n and p and decrease for h, the probability of the event $\{\widehat{\gamma}_n = \gamma_0^{(p)}\}$ tends to 1 under H_0. Hence $Q_n(\widehat{\gamma}_n)/\widehat{v}_n(\widehat{\gamma})$ behaves asymptotically like $Q_n(\gamma_0^{(p)})/\widehat{v}_n(\gamma_0^{(p)})$, even when p grows with the sample size. Therefore the test statistic we consider is

$$T_n = nh^{1/2} \frac{Q_n(\widehat{\gamma}_n)}{\widehat{v}_n(\widehat{\gamma}_n)} . \tag{5.3.3}$$

We will show that an asymptotic a-level test is given by $\mathbb{I}\left(T_n \geq z_{1-a}\right)$, where z_{1-a} is the $(1-a)$-th quantile of the standard normal distribution.

5.3.2 Behavior under the null hypothesis

In order to derive the asymptotic behavior of the statistic T_n under null hypothesis, below we introduce a set of assumptions on the data (Assumption D), and on the kernel and the rates of h and p (Assumption K).

Assumption D

(a) *The random vectors* $(U_1, X_1), \ldots, (U_n, X_n)$ *are independent draws from the random vector* $(U, X) \in L^2[0,1] \times L^2[0,1]$ *that satisfies* $\mathbb{E}\|U\|^8 < \infty$.

(b) *For any* $p \geq 1$ *and any* $\gamma \in \mathcal{S}^p$*, the d.f.* F_γ *is continuous.*

(c) $\exists\ \underline{\sigma}^2$, $C_1, C_2 > 0$ *and* $\nu > 2$ *such that :*

  (i) $0 < \underline{\sigma}^2 \leq \mathbb{E}(\langle U_1, U_2\rangle^2 \mathbb{I}_{\{\langle U_1, U_2\rangle \leq C_1\}} \mid X_1, X_2)$ *almost surely ;*

  (ii) $\mathbb{E}\left[\|U\|^\nu \mid X\right] \leq C_2$.

(d) *For any* $p \geq 1$, $\gamma_0^{(p)} \in B_p \subset \mathcal{S}^p$, B_p *are open subsets of* $\mathcal{S}^p$ *and* $B_p \times 0_{p'-p} \subset B_{p'}$, $\forall 1 \leq p < p'$ *where* $0_p \in \mathbb{R}^p$ *denotes the null vector of dimension* p.

A quick inspection of our proofs reveals that the key steps are derived conditionally on the X_i's. Thus the independence of the (U_i, X_i)'s could be relaxed to conditional independence of the U_i's given the X_i's to allow for some dependence among the X_i's. However, to reduce technicalities, we will work under the Assumption D-(a). The continuity condition required in Assumption D-(b) is a mild assumption that simplifies the NN smoothing. Assumption D-(c) will allow to prove that the variance of the statistics $Q_n(\gamma)$ is bounded away from zero and infinity uniformly with respect to γ. The very mild conditions imposed on B_p simplify the proofs for the consistency. These conditions are satisfied for instance when B_p is a half unit hypersphere.

Assumption K

(a) *The kernel* K *is a continuous density on real line such that* $K(x) = K(-x)$ *and* $K(\cdot)$ *is non increasing on* $[0, \infty)$*. Moreover the Fourier Transform of* K *is integrable.*

(b) $h \to 0$ *and* $nh^2 \to \infty$.

(c) $p \geq 1$ *increases to infinity with* n *and there exists a constant* $\lambda > 0$ *such that* $p \ln^{-\lambda} n$ *is bounded.*

The first step to derive a test statistic is the study of the behavior of the process $Q_n(\gamma)$, $\gamma \in B_p$, under H_0 when p is allowed to increase with the sample size. The following key lemma is crucially based on a powerful combinatorial result due to Cover (1967) on the number of possible orderings of $\langle X_1, \gamma\rangle, \cdots, \langle X_n, \gamma\rangle$ when γ belongs to the whole hypersphere $\mathcal{S}^p$, and on exponential inequalities for $U-$statistics.

Lemma 3.1 *Under Assumptions D and K and if H_0 holds true,*

$$\sup_{\gamma\in B_p\subset\mathcal{S}^p} |Q_n(\gamma)| = O_{\mathbb{P}}(n^{-1}h^{-1/2}p\ln n).$$

Moreover, if $\widehat{v}_n^2(\gamma)$ is the estimate defined in equation (5.3.1),

$$\sup_{\gamma\in B_p\subset\mathcal{S}^p} \{1/\widehat{v}_n^2(\gamma)\} = O_{\mathbb{P}}(1).$$

We now describe the behavior of $\widehat{\gamma}_n$ under H_0. A suitable rate α_n will make $\widehat{\gamma}_n$ to be equal to $\gamma_0^{(p)}$ with high probability. Under the null, α_n has to grow to infinity sufficiently fast to render the probability of the event $\{\widehat{\gamma}_n = \gamma_0^{(p)}\}$ close to 1. We will see below that, for better detection of alternative hypothesis, α_n should grow as slow as possible. Indeed, slower rates for α_n will allow the selection of directions $\hat{\gamma}_n$ that could be better suited than $\gamma_0^{(p)}$ for revealing the departure from the null hypothesis. The rate of p is also involved in the search of a trade-off for the rate of α_n : larger p renders slower the rate of uniform convergence to zero of $Q_n(\gamma)$, $\gamma \in B_p$, and hence requires larger α_n.

Lemma 3.2 *Under Assumptions D, K, for a positive sequence α_n, $n \geq 1$ such that $\alpha_n p^{-1}\ln^{-1} n \to \infty$,*

$$\mathbb{P}(\widehat{\gamma}_n = \gamma_0^{(p)}) \to 1, \qquad \textit{under } H_0.$$

The proof of Lemma 3.2 is similar to the proof of Lemma 3.2 in Lavergne and Patilea (2008) and hence will be omitted. The following result shows that the asymptotic critical values of our test statistic are standard normal.

Theorem 3.3 *Under the conditions of Lemma 3.2 and if the hypothesis H_0 in (5.1.1) holds true, the test statistic T_n converges in law to a standard normal. Consequently, the test given by $\mathbb{I}(T_n \geq z_{1-a})$, with z_a the $(1-a)-$quantile of the standard normal distribution, has asymptotic level a.*

Under suitable regularity conditions ensuring that the sample of U is estimated sufficiently accurate, the test statistic T_n will still have standard normal critical values. Patilea, Sánchez-Sellero and Saumard (2012) provide the complete arguments for their alternative test statistic in the case where the U_i's are the residuals of the functional linear model with scalar responses. Similar arguments could be adapted for the test considered in this paper. To keep this paper at reasonable length, the theoretical investigation of the extension to the case of estimated responses U_i will be omitted.

5.3.3 The behavior under the alternatives

Our test is consistent against the general alternative

$$H_1 : \mathbb{P}[\mathbb{E}(U \mid X) = 0] < 1,$$

that is the probability that the test statistic T_n is larger than any quantile z_{1-a} tends to one under H_1. This could be rapidly understood from the following simple inequalities :

$$\begin{aligned} T_n &= \frac{nh^{1/2}Q_n(\widehat{\gamma}_n)}{\widehat{v}_n(\widehat{\gamma}_n)} \\ &= \max_{\gamma\in B_p}\left\{nh^{1/2}Q_n(\gamma)/\widehat{v}_n(\gamma) - \alpha_n \mathbb{I}_{\{\gamma\neq\gamma_0^{(p)}\}}\right\} + \alpha_n \mathbb{I}_{\{\widehat{\gamma}_n\neq\gamma_0^{(p)}\}} \\ &\geq \max_{\gamma\in B_p}\frac{nh^{1/2}Q_n(\gamma)}{\widehat{v}_n(\gamma)} - \alpha_n \geq \frac{nh^{1/2}Q_n(\widetilde{\gamma})}{\widehat{v}_n(\widetilde{\gamma})} - \alpha_n, \quad \forall\widetilde{\gamma}\in B_p \subset \mathcal{S}^p, \end{aligned} \tag{5.3.4}$$

with $\widehat{v}_n(\gamma)$ defined in (5.3.1). Since $Var(\langle U_1, U_2\rangle \mid X_1, X_2) \geq \underline{\sigma}^2$, it is clear that $1/\widehat{v}_n(\widetilde{\gamma}) = O_{\mathbb{P}}(1)$ for all $\widetilde{\gamma}$. On the other hand, from Lemma 2.1, there exists p_0 and $\widetilde{\gamma}\in B_{p_0}$ such that the expectation of $Q_n(\widetilde{\gamma})$ stays away from zero as the sample size grows to infinity and h decrease to zero. On the other hand, for any $p > p_0$ and any n and h, clearly $\max_{\gamma\in B_p} Q_n(\gamma) \geq Q_n(\widetilde{\gamma})$, because $B_{p_0}\times 0_{p-p_0} \subset B_p$. All these facts show why our test is omnibus, that is consistent against nonparametric alternatives, provided that $p\to\infty$.

To state the consistency result, let $\delta(X)$ be some $L^2[0,1]$-valued function such that $\mathbb{E}[\delta(X)] = 0$ and $0 < \mathbb{E}[\|\delta(X)\|^4] < \infty$, and let r_n, $n\geq 1$ be sequence of real numbers that decrease to zero or $r_n = 1$, $\forall n$. Consider the sequence of alternative hypotheses

$$H_{1n} : U = U^0 + r_n\delta(X), \quad n\geq 1, \quad \text{with } U^0\in L^2[0,1],\ \mathbb{E}(U^0 \mid X) = 0.$$

We show below that such directional alternatives can be detected as soon as $r_n^2 nh^{1/2}/\alpha_n$ tends to infinity. This is exactly the condition one would obtain with scalar covariate ; see Lavergne and Patilea (2008). However, in the functional data framework, to obtain the convenient standard normal critical values, we need $1/\alpha_n = o(p^{-1}\ln^{-1} n)$. Hence, the rate r_n at which the alternatives H_{1n} tend to the null hypothesis should satisfy $r_n^2 nh^{1/2}/\{p\ln n\}\to\infty$.

Theorem 3.4 *Suppose that*

(a) Assumption D holds true with U replaced by U^0 ;

(b) Assumption K is satisfied and in addition $nh^4 \to \infty$ and there exists a constant C such that $|K(u) - K(v)| \leq C|u - v|$, $\forall u, v\in\mathbb{R}$;

(c) $\alpha_n/\{p \ln n\} \to \infty$ *and* r_n, $n \geq 1$ *is such that* $r_n^2 n h^{1/2}/\alpha_n \to \infty$ *;*

(d) $\mathbb{E}[\delta(X)] = 0$ *and* $0 < \mathbb{E}[\|\delta(X)\|^4] < \infty$;

(e) there exists p *and* $\widetilde{\gamma} \in B_p \subset \mathcal{S}^p$ *(independent of* n*) such that* $\mathbb{E}[\delta(X) \mid \langle X, \widetilde{\gamma} \rangle] \neq 0$ *and* $\forall t \in [0,1]$, *the Fourier Transform of* $\overline{\delta}(t,\cdot) = \mathbb{E}[\delta(X)(t) \mid F_\gamma(\langle X, \gamma \rangle) = \cdot]$ *is integrable ;*

Then the test based on T_n *is consistent against the sequence of alternatives* H_{1n}.

The additional Lipschitz condition on the kernel $K(\cdot)$ and the restriction on the bandwidth range in Theorem 3.4-(b) are reasonable technical conditions that simplifies the proof of the consistency. The zero mean condition for the function $\delta(\cdot)$ keeps U of zero mean under the alternative hypotheses H_{1n}. The existence of a vector $\widetilde{\gamma}$ such that $\mathbb{E}[\delta(X) \mid \langle X, \widetilde{\gamma} \rangle] \neq 0$ is guaranteed by Lemma 2.1-(B). In Theorem 3.4-(e) we impose a convenient mild technical condition on one of such vector.

5.3.4 Practical aspects

The goodness-of-fit procedure we propose in this paper requires the choice of several quantities : the orthonormal basis $\mathcal{R}$ in the space of X, the order p, the privileged direction $\gamma_0^{(p)}$ and the set B_p, the penalty amplitude α_n and the bandwidth h. In this section we provide some guidelines on how these quantities could be chosen by the practitioner. Before that, let us point out that the choice of the basis in the space of U is not really an issue. In applications, in what concerns the U_i's, the statistician only has to compute the $n(n-1)$ inner products $\langle U_i, U_j \rangle$ and this could be easily done with high accuracy in any basis the statistician would like to consider.

Our theoretical results above are derived for a fixed basis $\mathcal{R}$ in the space of X. The assumptions used to derive these results impose only very mild conditions on the basis $\mathcal{R}$, see Assumption D-(b) and condition (e) in Theorem 3.4. However, the choice of the basis could influence the finite sample performances of the test. Clearly, the practitioner would prefer a basis that allows for an accurate low-dimension representation of the covariate and hence for a low p in our testing procedure. A widely used basis is the one given by the eigenfunctions of the covariance operator Γ of X that is defined by

$$(\Gamma v)(t) = \int \sigma(t,s) v(s) ds, \qquad v \in L^2[0,1],$$

where X is supposed to satisfy the condition $\int \mathbb{E}(X^2(t))dt < \infty$ and $\sigma(t,s) = \mathbb{E}[\{X(t) - \mathbb{E}(X(t))\}\{X(s) - \mathbb{E}(X(s))\}]$ is supposed positive definite. Let $\lambda_1 \geq$

$\lambda_2 \geq \cdots$ denote the ordered eigenvalues of Γ and let $\mathcal{R} = \{\rho_1, \rho_2, \cdots\}$ be the corresponding basis of eigenfunctions of Γ that are usually called the functional principal components (FPC). The FPC represent the orthonormal basis of the Karhunen-Loève decomposition of X and provide optimal (with respect to the mean-squared error) low-dimension representations of X. See, for instance, Ramsay and Silverman (2005). In some cases where the law of X is given, the FPC are available. However, most of the time this is not the case and the FPC have to be estimated from the empirical covariance operator

$$(\widehat{\Gamma}v)(t) = \int \widehat{\sigma}(t,s)v(s)ds,$$

where $\widehat{\sigma}(t,s) = n^{-1}\sum_{i=1}^n \{X_i(t) - \overline{X}_n(t)\}\{X_i(s) - \overline{X}_n(s)\}]$ and $\overline{X}_n(t) = n^{-1}\sum_{i=1}^n X_i(t)$. Let $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots \geq 0$ denote the eigenvalues of $\widehat{\Gamma}$ and let $\widehat{\rho}_1, \widehat{\rho}_2, \cdots$ be the corresponding basis of eigenfunctions, that is the estimated FPC. We adopt the usual identification condition and we suppose that for any j, $\langle \rho_j, \widehat{\rho}_j \rangle \geq 0$. For any $\gamma = (\gamma_1, \cdots, \gamma_p) \in \mathcal{S}^p$ let us define

$$\langle X_i, \gamma \rangle_n = \sum_{k=1}^p \gamma_k \int_{[0,1]} X_i(t)\widehat{\rho}_k(t)dt.$$

Let $\widehat{T}_n$ be the test statistic obtained from equations (5.3.2) and (5.3.3) after replacing all the inner products $\langle X_i, \gamma \rangle$ by the estimated versions $\langle X_i, \gamma \rangle_n$. Below we show that the test $\mathbb{I}(\widehat{T}_n \geq z_{1-a})$ behaves asymptotically like the test $\mathbb{I}\left(T_n \geq z_{1-a}\right)$. For the behavior under the null hypothesis, no additional assumption is required. For the consistency we impose mild conditions on X and a slightly more restrictive bandwidth range.

Corollary 3.5 *a) Under the same conditions, the conclusion of Theorem 3.3 remains true if T_n is replaced by $\widehat{T}_n$.*

b) In addition to the conditions of Theorem 3.4 assume that

1. *there exist $C, \eta > 0$ such that $\lambda_j - \lambda_{j+1} \geq Cj^{-\eta}$, $\forall j \geq 1$;*
2. *there exists $\varrho > 0$ such $\mathbb{E}[\exp(\varrho\|X\|)] < \infty$;*
3. *the vector $\widetilde{\gamma} \in\subset \mathcal{S}^p$ in condition (e) of Theorem 3.4 is such the variable $\langle X, \widetilde{\gamma} \rangle$ has a bounded density $f_{\widetilde{\gamma}}$;*
4. *$nh^4/p^{\eta+1}\ln^2 n \to \infty$.*

Then the conclusion of Theorem 3.4 remains true if T_n is replaced by $\widehat{T}_n$.

The first condition on the spacings between the ordered eigenvalues of Γ is a mild version of a common condition in functional data modeling. The exponential moment condition on $\|X\|$ is satisfied in many situations. For instance, if X is a gaussian process, it suffices to have some $\alpha > 1$ and $c > 0$ such that $\lambda_j \leq cj^{-\alpha}$ for all $j \geq 1$. Moreover, let us recall that the null hypothesis does not change if the covariate X is transformed by a one-to-one map. Such a transformation could be chosen such that the exponential moment condition is fulfilled. Then, the FPC basis could be estimated from the transformed X and our Corollary 3.5 applies. The exponential moment condition on $\|X\|$ could be relaxed at the expense of more restrictions on the bandwidth range. Finally, in view of Lemma 2.1-(B), almost any unit norm vector of finite but sufficiently large dimension is a candidate to be $\widetilde{\gamma}$. Hence the bounded density condition for some $\langle X, \widetilde{\gamma}\rangle$ is also a mild restriction. For instance, it is satisfied for any unit norm vector $\widetilde{\gamma} \in \mathbb{R}^p$ if X is a gaussian process.

We proved in the previous sections that the choice of $\gamma_0^{(p)}$ is irrelevant for the theory. However, in practice the choice of $\gamma_0^{(p)}$ could be related to prior information of the practitioner on a class of alternatives. Concerning the set B_p, since $Q_n(\gamma) = Q_n(-\gamma)$ for any $\gamma \in \mathcal{S}^p$, one could restrict the set B_p to a half unit hypersphere like $\{\gamma \in \mathcal{S}^p : \gamma_1 \geq 0\}$. One could restrict B_p even more, and hence to speed optimization algorithms, when some prior information indicates a set of directions that would be able to detect alternatives.

Under the null hypothesis, if $n \to \infty$ and p increases with n at a suitable rate, the ratio $nh^{1/2}Q_n(\gamma)/\widehat{v}_n(\gamma)$ behaves like a standard normal for any given sequence of $\gamma \in \mathcal{S}^p$. Meanwhile the supremum of this ratio with respect to $\gamma \in B_p$ diverges in probability with a rate smaller or equal to $p \ln n$. Hence α_n has to grow to infinity faster than $p \ln n$. In practice, larger α_n will likely result in taking $\widehat{\gamma}_n = \gamma_0^{(p)}$ and in this case the standard normal critical values will be accurate even for moderate samples sizes. Having $\widehat{\gamma}_n = \gamma_0^{(p)}$ might be reasonable when the practitioner judges $\gamma_0^{(p)}$ trustful for detecting alternatives. On contrary, smaller α_n will likely lead to a value of the test statistic equal to the maximal value of $nh^{1/2}Q_n(\gamma)/\widehat{v}_n(\gamma)$ and hence in general the test will be too liberal. Meanwhile, smaller α_n is preferable for detecting general alternatives. On the basis of our detailed simulation investigations, we recommend smaller values for α_n and a correction of the critical values through resampling, as explained below.

Next, let us discuss the influence of the bandwidth choice. In our theory the bandwidth choice does not appear in the asymptotic approximation of the size of the test. However, with finite samples, a size correction is often necessary. We propose to do this correction using a simple wild bootstrap procedure that we describe below. Alternatively, one could look for more elaborate high-order approximations of the size function, as for instance those considered by Gao and Gijbels (2008) ;

see also the reference therein. Such a theoretical investigation could likely be reproduced in our framework in a case of a fixed, finite-dimension $\gamma_0^{(p)}$ and using a suitable control of α_n, but would be much more involved in a general setting where $p \uparrow \infty$. Concerning the power oriented choice of the bandwidth, again the high-order asymptotics of Gao and Gijbels (2008) could be reconsidered with some particular alternatives, but the general case seems to be much more complicated. A perhaps more direct approach for optimal (with respect to the power) bandwidth choice, would be to follow the idea of Horowitz and Spokoiny (2001) and to define a test statistic as the maximum of T_n over a grid of bandwidths ; see also Guerre and Lavergne (2005). In view of Theorem 3.5 of Lavergne and Patilea (2008), such a procedure would have, under suitable technical conditions, some optimality property uniformly over smoothness classes of functions (e.g., Hölder classes), provided such functions depend only on some $\langle X, \gamma \rangle$ for a fixed, finite-dimension γ. At the price of some technical assumptions, one could expect that the results extend to single-index functions depending on some $\langle X, \beta \rangle$ for some fixed $\beta \in L^2[0,1]$. All these challenging bandwidth choice aspects deserve a separate careful investigation and will be considered in future work.

Before ending this section let us propose a wild procedure that could be used for correcting the finite sample critical values. The bootstrap sample, denoted by U_i^b , $1 \leq i \leq n$, is obtained as follows : $U_i^b = \zeta_i U_i$, $1 \leq i \leq n$, where ζ_i, $1 \leq i \leq n$ are independent random variables following the two-points distribution proposed by Mammen (1993), that is, $\zeta_i = -(\sqrt{5}-1)/2$ with probability $(\sqrt{5}+1)/(2\sqrt{5})$ and $\zeta_i = (\sqrt{5}+1)/2$ with probability $(\sqrt{5}-1)/(2\sqrt{5})$. A bootstrap test statistic T_n^b is built from a bootstrap sample as was the original test statistic. Similarly, let $\widehat{T}_n^b$ be the bootstrap test statistic obtained by from this procedure applied with the estimated FPC basis. When this scheme is repeated many times, the bootstrap critical value $z^\star_{1-a,n}$ at level a is the empirical $(1-a)-$th quantile of the bootstrapped test statistics. This critical value is then compared to the initial test statistic. The asymptotic validity of this bootstrap procedure is guaranteed by the following result.

Theorem 3.6 *Suppose that the conditions of Theorem 3.3 hold. Then, under* H_0

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(T_n^b \leq x \mid U_1, X_1, \cdots, U_n, X_n\right) - \mathbb{P}(T_n \leq x) \right| \to 0, \qquad \textit{in probability.}$$

The same statement remains true with $\widehat{T}_n^b$ *and* $\widehat{T}_n$ *replacing* T_n^b *and* T_n*, respectively.*

5.4 Empirical study

A simulation study was carried out to assess the behavior of the proposed methods under the null and with different types of effects under the alternative. For comparison with the procedure proposed by Kokoszka *et al.* (2008), we considered a sample size $n = 40$. The critical values of our procedure were approximated by a wild bootstrap procedure as described above.

5.4.1 Simulation study

The first situation we considered was a functional linear model, given by

$$U_i(t) = \int_0^1 \psi(s,t) X_i(s)\, ds + \epsilon_i(t), \qquad 1 \le i \le n$$

where X_i and ϵ_i are independent Brownian bridges and ψ is square-integrable over $[0,1) \times [0,1)$. The kernel ψ was chosen to be $\psi(s,t) = c \cdot \exp(t^2 + s^2)/2$, with $c = 0$ under the null and $c = 0.3$ under the alternative.

The well-known Karhunen-Loeve decomposition of the Brownian bridge provides a good approximation of the covariate function. Thus, the orthonormal basis of eigenfunctions $\mathcal{R} = \{\sqrt{2}\sin(j\pi t) : 0 \le t \le 1, j = 1, 2 \ldots\}$ seems a good choice for our test statistic. Different possibilities for the privileged direction $\gamma_0^{(p)}$ were considered. The direction $\gamma_0^{(p)} = (1, 0, \ldots, 0) \in \mathcal{S}^p$ generally provides a powerful test. Here we present the results for an uninformative direction, with the same coefficients in all basic elements. For the penalization we used the value $\alpha_n = 1$, which provides a good trade-off between the privileged direction and the direction maximizing the standardized statistic.

To compute the statistic for each direction, we used the Epanechnikov kernel, $K(x) = (1 - x^2)\mathbb{I}_{\{|x| \le 1\}}$. A grid of bandwidths was considered in order to explore the effect of the bandwidth on the power of the test.

The number of basic components was $p = 3$. For the optimization in the hypersphere $\mathcal{S}^p$, a grid of 1200 points was used. For each original sample, we used 499 bootstrap samples to compute the critical value. One thousand original samples of size $n = 40$ were generated to approximate the percentages of rejection.

Figure 5.1 shows the empirical powers obtained for a grid of values of the bandwidth both under the null hypothesis of no-effect and under the functional linear alternative. We observe that the power is not very much affected by the bandwidth around a possibly optimal value. For purposes of comparison, the empirical power of the Kokoszka et al. (2008)'s test is also shown. These authors proposed a test of the functional linear effect, that is, a test specially designed to detect the alternative of a functional linear effect versus the no-effect. Our test provides similar or even better power than the Kokoszka *et al.*'s parametric test in their ideal framework. The level is quite well respected for any of the considered bandwidths.

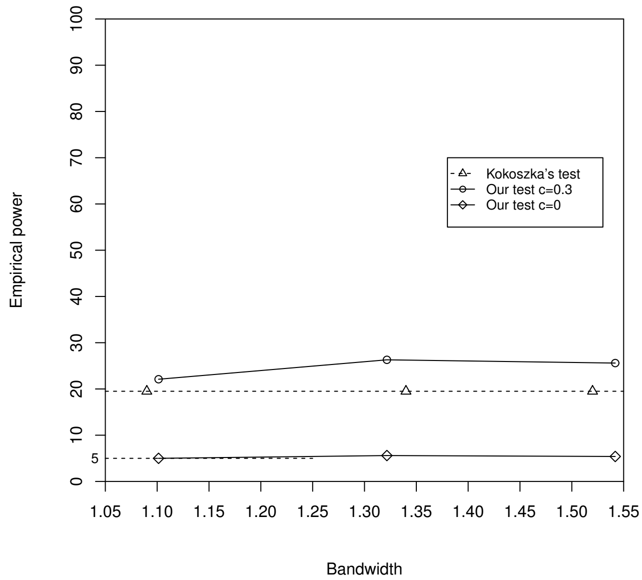

FIGURE 5.1 – Testing the null-effect versus a functional linear alternative.

Another alternative was considered of the following type :

$$U_i(t) = \beta(t)X_i(t) + \epsilon_i(t), \qquad 1 \leq i \leq n$$

where X_i and ϵ_i are independent Brownian bridges (as in the previous situation) and β is a square-integrable function on $[0, 1]$. This is the so-called concurrent model studied in detail in Ramsay and Silverman (2005), where the covariate at time t, $X_i(t)$, only influences the response function at time t, $U_i(t)$. The function β was $\beta(t) = c \cdot \exp(-4(t - 0.3)^2)$, with $c = 0$ under the null and $c = 0.6$ under the alternative.

Figure 5.2 shows the power of our test under the concurrent alternative, in comparison with Kokoszka *et al.*'s test. In this case, Kokoszka *et al.*'s test is slightly more powerful than ours. This is not necessarily surprising since the concurrent model is in a sense a degenerate functional linear model.

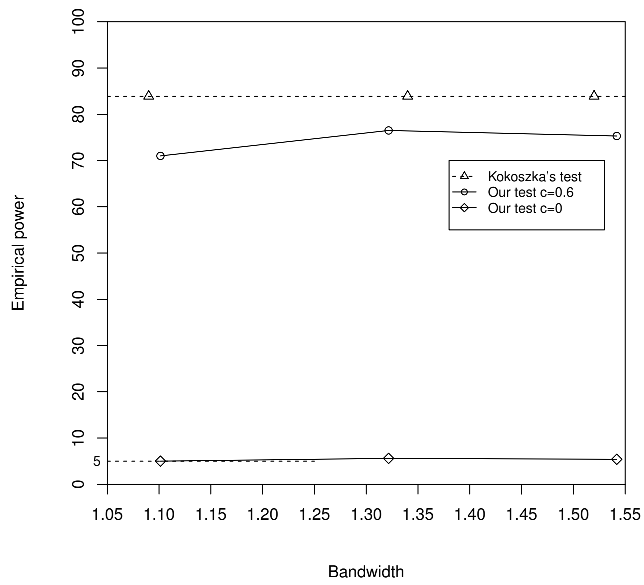

FIGURE 5.2 – Testing the null-effect versus a concurrent model alternative.

A completely nonlinear alternative was also considered. In this case a quadratic model of this type was generated :

$$U_i(t) = H\left(X_i(t)\right) + \epsilon_i(t), \qquad 1 \leq i \leq n$$

where X_i and ϵ_i are independent Brownian motion and Brownian bridge, respectively, and $H_2(x) = x^2 - 1$ Since the covariate function is a Brownian motion, instead of the Brownian bridge of the previous situations, the basis was chosen as the orthonormal basis of eigenfunctions of the Brownian motion.

Figure 5.3 shows the percentages of rejections under the null and under this quadratic alternative for a range of bandwidths. The power of the Kokoszka *et al.*'s test is also plotted. As expected, Kokozska *et al.*'s test, which was designed to detect only linear effects, is not powerful under this quadratic alternative.

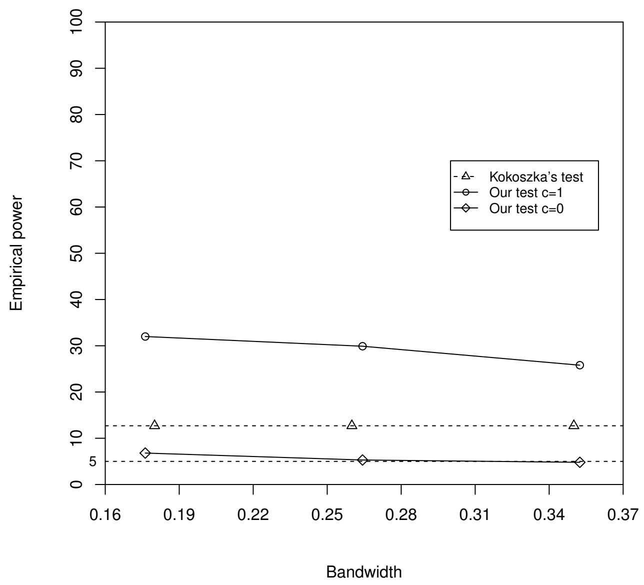

FIGURE 5.3 – Testing the null-effect versus a quadratic alternative.

5.5 Appendix : technical proofs

In this section $c, c_1, C, C_1, ...$ denote constants that may have different values from line to line. Moreover, for any integrable function ϕ defined on the real line, $\mathcal{F}[\phi]$ denotes its Fourier Transform, that is $\mathcal{F}[\phi](t) = \int_{\mathbb{R}} \phi(x) \exp\{-2\pi itx\}dx$. Finally, recall that if $X = \sum_{j=1}^{\infty} x_j\rho_j$, then $X^{(p)} = \sum_{j=1}^{p} x_j\rho_j$.

Proof of Lemma 2.1. (A) We have

$$
\begin{aligned}
\mathbb{E}(U \mid X) = 0 &\Leftrightarrow \mathbb{E}(\langle U, \rho_j\rangle \mid X) = 0,\ \forall j \geq 1\\
&\Leftrightarrow \mathbb{E}(\langle U, \rho_j\rangle \mid X^{(p)}) = 0,\ \forall j \geq 1, \forall p \geq 1\\
&\Leftrightarrow \mathbb{E}(\langle U, \rho_j\rangle \mid \langle X, \gamma\rangle) = 0,\ \forall j \geq 1, \forall p \geq 1, \forall \gamma \in \mathcal{S}^p\\
&\Leftrightarrow \mathbb{E}(U \mid \langle X, \gamma\rangle) = 0,\ \forall p \geq 1, \forall \gamma \in \mathcal{S}^p\\
&\Leftrightarrow \mathbb{E}(U \mid F_\gamma(\langle X, \gamma\rangle)) = 0,\ \forall p \geq 1, \forall \gamma \in \mathcal{S}^p
\end{aligned}
$$

The first and the fourth equivalence in the last display are due to the fact that $\mathcal{R}$ is a basis in $L^2[0,1]$. Next, note that by Cauchy-Schwarz inequality $\forall j$, $\mathbb{E}|\langle U, \rho_j\rangle| \leq \mathbb{E}\|U\| < \infty$. Thus the second equivalence in the last display is guaranteed elementary properties of the conditional expectations and the Doob's Martingale Convergence Theorem, while the third equivalence is given by Lemma 2.1-(A) of Lavergne and Patilea (2008). For the last equivalence recall that for any random variable Y with d.f. F, $\mathbb{P}(F^{-1} \circ F(Y) \neq Y) = 0$ where $F^{-1}(t) = \{y : F(y) \geq t\}, \forall 0 < t < 1$; see for instance Proposition 3, Chapter 1 in Shorack and Wellner (1986). Deduce that $\mathbb{E}(U \mid \langle X, \gamma \rangle) = \mathbb{E}(U \mid F_\gamma(\langle X, \gamma \rangle))$. To complete the proof of part (A) it suffices to note that

$$\begin{aligned}
\mathbb{E}\left[\langle U, \mathbb{E}\left(U \mid \langle X, \gamma \rangle\right)\rangle\right] &= \mathbb{E}\left[\|\mathbb{E}(U \mid \langle X, \gamma \rangle)\|^2\right] \\
&= \mathbb{E}\left[\|\mathbb{E}(U \mid F_\gamma(\langle X, \gamma \rangle))\|^2\right] \\
&= \mathbb{E}\left[\langle U, \mathbb{E}\left\{U \mid F_\gamma(\langle X, \gamma \rangle)\right\}\rangle\right].
\end{aligned}$$

(B) First note that $\mathcal{A} \subset \bigcap_{j \geq 1} \mathcal{A}_j$ where

$$\mathcal{A}_j = \{\gamma \in \mathcal{S}^p : \mathbb{E}(\langle U, \rho_j \rangle \mid \langle X, \gamma \rangle) = 0 \ a.s.\}.$$

Now, if $\mathbb{P}[\mathbb{E}(U \mid X) = 0] < 1$, then there exists $j \geq 1$ such that $\mathbb{P}[\mathbb{E}(\langle U, \rho_j \rangle \mid X) = 0] < 1$. For any arbitrarily fixed $j \geq 1$, Lemma 2.1 in Patilea, Saumard and Sanchez (2012) allows to deduce that there exists $p_0 \geq 1$ such that, for any $p \geq p_0$, $\mathcal{A}_j$ has Lebesgue measure zero on $\mathcal{S}^p$ and is not dense. Since $\mathcal{A}$ is included in any $\mathcal{A}_j$, the conclusion follows. ∎

Lemma 5.1 *Let K be a density satisfying Assumption K-(a) and assume that $h \to 0$ and $nh \to \infty$. Let*

$$S_{ni} = \frac{1}{(n-1)h} \sum_{1 \leq j \leq n,\, i \neq j} K\left(\frac{i-j}{nh}\right) \quad \text{and} \quad S_n = \frac{1}{n} \sum_{1 \leq i \leq n} S_{ni}.$$

Then exists constants c_1, c_2 such that for sufficiently large n

$$0 < c_1 \leq \min_{1 \leq i \leq n} S_{ni} \leq \max_{1 \leq i \leq n} S_{ni} \leq c_2 < \infty.$$

Moreover, $S_n \to 1$.

Proof. Clearly that $S_n - \widetilde{S}_n \to 0$ where

$$\widetilde{S}_n = \frac{1}{n^2 h} \sum_{1 \leq i,j \leq n} K\left(\frac{i-j}{nh}\right).$$

If $[a]$ denote the integer part of any real number a, we can write

$$\begin{aligned}
\widetilde{S}_n &= \int_{1/n}^{(n+1)/n} \int_{1/n}^{(n+1)/n} h^{-1} K\left(\frac{[ns]-[nt]}{nh}\right) dsdt \\
&= \int_{1/n}^{(n+1)/n} \int_{1/nh-t/h}^{1/h+1/nh-t/h} K\left(\frac{[nt+nzh]-[nt]}{nh}\right) dzdt \qquad [z=(s-t)/h] \\
&= \int_{1/n}^{(n+1)/n} \int_{1/nh-t/h}^{1/h+1/nh-t/h} K(z)\, dzdt + o(1) \\
&= \int_{-1/h}^{1/h} \int_{1/n-zh}^{1+1/n-zh} dt K(z)\, dz + o(1) \qquad [\text{Fubini}] \\
&\to 1,
\end{aligned}$$

where the order $o(1)$ of the reminder on the right-hand side of the third equality could be obtained as a consequence of the fact K is symmetric and monotonic. Hence $S_n \to 1$. Similarly, we can write

$$\begin{aligned}
\widetilde{S}_{ni} &= \int_{1/n}^{(n+1)/n} h^{-1} K\left(\frac{i-[nt]}{nh}\right) dt \\
&= \int_{(1-i)/nh}^{1/h+(1-i)/nh} K\left(\frac{i-[i+nzh]}{nh}\right) dz \qquad [z=(t-i/n)/h] \\
&= \int_{(1-i)/nh}^{1/h+(1-i)/nh} K(z)\, dz + o(1).
\end{aligned}$$

Deduce that

$$\int_0^1 K(z)dz + \underline{r}_{ni} \leq \widetilde{S}_{ni} \leq \int_{\mathbb{R}} K(z)dz + \overline{r}_{ni}$$

where $\max_{1\leq i\leq n}\{|\underline{r}_{ni}| + |\overline{r}_{ni}|\} = o(1)$. The result follows. ■

One of the ingredients we will use for the proof of Lemma 3.1 is a moment inequality for $U-$statistics presented in Lemma 5.2 below and due to Giné, Latała and Zinn (2000). To state the result we will use, let us introduce some notation. Let $Z_1, \cdots, Z_n$ be independent random variables (not necessarily with the same distribution) taking values in a measurable space $(\mathcal{Z}, \Upsilon)$. Let $h_{i,j}(\cdot,\cdot)$, $1 \leq i,j \leq n$ be real-valued measurable functions on $\mathcal{Z}^2$ such that $h_{i,j}(z_i, z_j) = h_{j,i}(z_j, z_i)$ and $\mathbb{E}[h_{i,j}(z_i, Z_j)] = 0$, $\forall 1 \leq i,j \leq n$, $\forall z_i, z_j$. The functions $h_{i,j}$ could be different for different values of n. Define

$$A_n = \max_{i,j} \|h_{i,j}(\cdot,\cdot)\|_\infty, \quad B_n^2 = \max_j \left\| \sum_i \mathbb{E}[h_{i,j}^2(Z_i,\cdot)] \right\|_\infty, \quad C_n^2 = \sum_{i,j} \mathbb{E}[h_{i,j}^2(Z_i, Z_j)], \tag{5.5.1}$$

and

$$D_n = \sup\left\{ \mathbb{E}\sum_{i,j} h_{i,j}(Z_i, Z_j) f_i(Z_i) g_j(Z_j) \;:\; \mathbb{E}\sum_i f_i^2(Z_i) \le 1,\ \mathbb{E}\sum_j g_j^2(Z_j) \le 1 \right\}. \tag{5.5.2}$$

The following result is simplified version of Theorem 3.3 in Giné, Latała and Zinn (2000).

Lemma 5.2 *There exist an universal constant $L < \infty$ (in particular, independent on n and the functions $h_{i,j}$) such that*

$$\mathbb{P}\left\{ \sum_{1\le i\ne j\le n} h_{i,j}(Z_i, Z_j) \ge t \right\} \le L \exp\left[-\frac{1}{L} \min\left(\frac{t^2}{C_n^2}, \frac{t}{D_n}, \frac{t^{2/3}}{B_n^{2/3}}, \frac{t^{1/2}}{A_n^{1/2}} \right) \right], \qquad \forall t > 0.$$

Let $\gamma \in \mathcal{S}^p$ and let $x_1, \cdots, x_n$ be an arbitrary collection of non-random points in $L^2[0,1]$. Consider $\widetilde{Z}_1, \cdots, \widetilde{Z}_n$ independent random variables with values in $L^2[0,1]$ such that for each $1 \le i \le n$ the law of $\widetilde{Z}_i$ is the conditional law of U_i given $X_i = x_i$. We will apply Lemma 5.2 with $h_{i,i} \equiv 0$ and for $1 \le i \ne j \le n$

$$h_{i,j}(Z_i, Z_j) = \frac{\langle Z_i\,, Z_j\rangle}{n(n-1)hM^2} K_h\left(F_{\gamma,n}(\langle x_i, \gamma\rangle) - F_{\gamma,n}(\langle x_j, \gamma\rangle) \right), \tag{5.5.3}$$

where $Z_i = \widetilde{Z}_i \mathbb{I}_{\{\|\widetilde{Z}_i\| \le M\}} - \mathbb{E}[\widetilde{Z}_i \mathbb{I}_{\{\|\widetilde{Z}_i\| \le M\}}]$, $M > 0$ is some constant (that we will allow to increase with n).[§] Here $F_{\gamma,n}$ is the empirical d.f. of the sample $\langle x_1, \gamma\rangle, \cdots, \langle x_n, \gamma\rangle$. The functions $h_{i,j}(\cdot,\cdot)$ vanish outside the rectangle $[-2M, 2M] \times [-2M, 2M]$. The following lemma provides upper bounds for the quantities A_n to D_n in this setup. The bounds are independent of the collection $x_1, \cdots, x_n \in L^2[0,1]$, and of $p \ge 1$ and $\gamma \in \mathcal{S}^p$.

Lemma 5.3 *Under the conditions of Lemma 3.1, for $h_{i,j}$ defined as in (5.5.3)*

$$A_n = \frac{\|K\|_\infty}{n(n-1)h}, \quad B_n^2 \le \frac{c}{n^3 h M^2}, \quad C_n^2 \le \frac{c}{n^2 h M^4} \quad \text{and} \quad D_n \le \frac{c}{nM^2},$$

for some constant c depending only on the upper bound of $\mathbb{E}(\|U\|^2 \mid X)$ and $\int K^2$.

Proof. The bound for A_n is obvious. For C_n^2 note that

$$\mathbb{E}[h_{i,j}^2(Z_i, Z_j)] = \frac{M^{-4}}{n^2(n-1)^2 h} \mathbb{E}\left\{ \mathbb{E}\left[\langle Z_i\,, Z_j\rangle^2\right] h^{-1} K_h\left(F_{\gamma,n}(\langle x_i, \gamma\rangle) - F_{\gamma,n}(\langle x_j, \gamma\rangle) \right) \right\}.$$

§. Note that in particular the $\mathbb{E}[Z_i \mathbb{I}_{\{\|Z_i\| \le M\}}]$ coincide with the values $\mathbb{E}[U_i \mathbb{I}_{\{\|U_i\| \le M\}} \mid X_i = x_i]$.

By Cauchy-Schwarz inequality and triangle inequality and recalling that $\widetilde{Z}_i$ is distributed according to the conditional law of U_i given $X_i = x_i$,

$$\mathbb{E}\left[\langle Z_i\ , Z_j\rangle^2\right] \leq 16\mathbb{E}\left[\|\widetilde{Z}_i\|^2\right]\mathbb{E}\left[\|\widetilde{Z}_j\|^2\right] \leq 16C^2,$$

for any constant C that bounds from above $\mathbb{E}(\|U\|^2 \mid X)$, see Assumption D-(c). Finally, note that

$$\frac{1}{n(n-1)h}\sum_{1\leq i\neq j\leq n} K_h\left(F_{\gamma,n}(\langle x_i,\gamma\rangle) - F_{\gamma,n}(\langle x_j,\gamma\rangle)\right) = \frac{1}{n(n-1)h}\sum_{1\leq i\neq j\leq n} K\left(\frac{i-j}{nh}\right)$$

and apply the second part of Lemma 5.1 to derive the bound for C_n^2. To derive the bound for B_n^2 recall that $h_{i,j}(Z_j, z)$ vanishes for $|z| > 2M$, use again Cauchy-Schwarz inequality and triangle inequality and the first part of Lemma 5.1. For the bound of D_n, using Cauchy-Schwarz inequality and the independence of Z_i and Z_j, we can write

$$\begin{aligned}
\mathbb{E}\sum_{i,j} h_{i,j}(Z_i,Z_j)f_i(Z_i)g_j(Z_j) &\leq \sum_{i,j}\frac{\mathbb{E}|\langle Z_i f_i(Z_i), Z_j g_j(Z_j)\rangle|}{n(n-1)hM^2}K_h(F_{\gamma,n}(\langle x_i,\gamma\rangle) - F_{\gamma,n}(\langle x_j,\gamma\rangle))\\
&\leq \sum_{i,j}\frac{16C^2\mathbb{E}|f_i(Z_i)|\mathbb{E}|g_j(Z_j)|}{n(n-1)hM^2}K_h(F_{\gamma,n}(\langle x_i,\gamma\rangle) - F_{\gamma,n}(\langle x_j,\gamma\rangle))\\
&\leq \frac{16C^2}{M^2}|||\mathcal{K}|||_2,
\end{aligned}$$

where C is such that $\mathbb{E}(\|U\|^2 \mid X) \leq C$ and $\mathcal{K}$ is the matrix with elements

$$\mathcal{K}_{ij} = K\left((i-j)/nh\right)/[n(n-1)h], \quad i \neq j, \quad \text{and} \quad \mathcal{K}_{ii} = 0, \tag{5.5.4}$$

and $|||\mathcal{K}|||_2$ is the spectral norm of $\mathcal{K}$. By definition, $|||\mathcal{K}|||_2 = \sup_{u\in\mathbb{R}^n, u\neq 0}\|\mathcal{K}u\|/\|u\|$ and $|u'\mathcal{K}w| \leq |||\mathcal{K}|||_2\|u\|\|w\|$ for any $u, w \in \mathbb{R}^n$. By Lemma 5.1, for any $u \in \mathbb{R}^n$,

$$\begin{aligned}
\|\mathcal{K}u\|^2 &= \sum_{i=1}^n\left(\sum_{j=1,j\neq i}^n \frac{K_h\left((i-j)/nh\right)}{h\,n(n-1)}u_j\right)^2\\
&\leq \sum_{i=1}^n\left(\sum_{j=1,j\neq i}^n \frac{K_h\left((i-j)/nh\right)}{h\,n(n-1)}\right)\sum_{j=1,j\neq i}^n\frac{K_h\left((i-j)/nh\right)}{h\,n(n-1)}u_j^2\\
&\leq \|u\|^2\left[\max_{1\leq i\leq n}\left(\sum_{j=1,j\neq i}^n\frac{K_h\left((i-j)/nh\right)}{h\,n(n-1)}\right)\right]^2\\
&\leq cn^{-2}\|u\|^2,
\end{aligned} \tag{5.5.5}$$

for some constant $c > 0$. The bound for D_n follows immediately. ■

Another ingredient is an upper bound for the number of different possible orderings in the sample $\langle X_1, \gamma\rangle, \cdots, \langle X_n, \gamma\rangle$ when γ belongs to the unit hypersphere in $\mathbb{R}^p$ (obviously the same number is obtained if γ is allowed to belong to the whole space $\mathbb{R}^p$). Let $w_1, \cdots, w_n$ be a collection of n points in $\mathbb{R}^p$ and let π be a permutation of the set of integers $\{1, 2, \cdots, n\}$. Following Cover (1967), we shall say that $\gamma \in \mathcal{S}^p$ induces the ordering π if

$$\langle w_{\pi(1)}, \gamma\rangle < \langle w_{\pi(2)}, \gamma\rangle < \cdots < \langle w_{\pi(n)}, \gamma\rangle.$$

Conversely, the ordering π will be said to be linearly inducible if there exists such vector γ. The following result is due to Cover (1967).

Lemma 5.4 *There are precisely $q(n, p)$ linearly inducible orderings of n points in general position in $\mathbb{R}^p$, where*

$$q(n,p) = 2\sum_{k=0}^{p-1} S_{n,k} = 2\left[1 + \sum_{2\leq i\leq n-1} i + \sum_{2\leq i<j\leq n-1} ij + \cdots\right] \quad (p\ \textit{terms}),$$

where $S_{n,k}$ is the number of the $(n-2)!/(n-2-k)!k!$ possible products of numbers taken k at a time without repetition from the set $\{2, 3, \cdots, n-1\}$

By Lemma 5.4 we obtain a simple upper bound for $q(n, p)$ when $n \geq 2p$, that is

$$q(n,p) \leq 2[1 + n^2 + \cdots + n^p] \leq n^{p+1}. \tag{5.5.6}$$

Proof of Lemma 3.1. Fix M that depends on n in a way that will be specified below. Let

$$Q_{M,n}(\gamma) = \frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} \langle U_{M,i}, U_{M,j}\rangle \frac{1}{h} K_h\left(F_{\gamma,n}(\langle X_i,\gamma\rangle) - F_{\gamma,n}(\langle X_j,\gamma\rangle)\right), \quad \gamma \in \mathcal{S}^p,$$

where $U_{M,i} = U_i \mathbb{I}_{\{\|U_i\|\leq M\}} - \mathbb{E}[U_i \mathbb{I}_{\{\|U_i\|\leq M\}}]$. We can write

$$\mathbb{P}\left(\sup_{\gamma\in\mathcal{S}^p} |Q_{M,n}(\gamma)| > \frac{tp\ln n}{nh^{1/2}}\right) = \mathbb{E}\left[\mathbb{P}\left(\sup_{\gamma\in\mathcal{S}^p} |Q_{M,n}(\gamma)| > \frac{tp\ln n}{nh^{1/2}} \mid X_1, \cdots, X_n\right)\right]$$

In view of Lemma 5.4, for any n, p, given $X_1, \cdots, X_n$ there exists a set $\mathcal{O}_{np} \subset \mathbb{R}^p$ with at most n^p elements, that depend on $X_1, \cdots, X_n$, such that

$$\sup_{\gamma\in\mathcal{S}^p} |Q_{M,n}(\gamma)| = \sup_{\gamma\in\mathcal{O}_{np}} |Q_{M,n}(\gamma)|.$$

Let $b_n = M^{-2}n^{-1}h^{-1/2}p\ln n$. By Lemmas 5.2 and 5.4 deduce that there exists an universal constant L such that for any $t > 0$,

$$\mathbb{P}\Bigg(\sup_{\gamma\in\mathcal{S}^p}|Q_{M,n}(\gamma)| > \frac{tp\ln n}{nh^{1/2}} \mid X_1,\cdots,X_n\Bigg) \leq \sum_{\gamma\in\mathcal{O}_{np}}\mathbb{P}\big(|M^{-2}Q_{M,n}(\gamma)| > tb_n \mid X_1,\cdots,X_n\big)$$
$$\leq \max\{L,1\}\exp\left[(p+1)\ln n - \frac{1}{L}\min\left(\frac{(tb_n)^2}{C_n^2},\frac{tb_n}{D_n},\frac{(tb_n)^{2/3}}{B_n^{2/3}},\frac{(tb_n)^{1/2}}{A_n^{1/2}}\right)\right].$$

Now, take $M = n^{1/4-a}$ for some (small) $a > 0$ and notice that the exponential bound in the last display is independent of $X_1,\cdots,X_n$ and tends to zero for any t. Deduce that

$$\sup_{\gamma\in\mathcal{S}^p}|Q_{M,n}(\gamma)| = O_{\mathbb{P}}(n^{-1}h^{-1/2}p\ln n).$$

Next we show that $\sup_{\gamma\in\mathcal{S}^p}|Q_n(\gamma) - Q_{M,n}(\gamma)| = o_{\mathbb{P}}(n^{-1}h^{-1/2}p\ln n)$. Let

$$R_{1n}(\gamma) = \frac{1}{n(n-1)}\sum_{1\leq i\neq j\leq n}\langle U_{M,i}, U_j - U_{M,j}\rangle\frac{1}{h}K_h\left(F_{\gamma,n}(\langle X_i,\gamma\rangle) - F_{\gamma,n}(\langle X_j,\gamma\rangle)\right), \quad \gamma\in\mathcal{S}^p,$$

and $R_{2n}(\gamma) = Q_n(\gamma) - Q_{M,n}(\gamma) - 2R_{1n}(\gamma)$. Since $K(\cdot)$ is bounded, we have

$$\mathbb{E}\left[\sup_{\gamma}|R_{1n}(\gamma)| \mid X_1,\cdots,X_n\right] \leq Ch^{-1}\mathbb{E}\left(\|U_{M,i}\|\|U_j - U_{M,j}\|\right)$$
$$\leq 2Ch^{-1}\mathbb{E}\left(\|U_i\|\right)\mathbb{E}\left(\|U_j - U_{M,j}\|\right).$$

By Hölder inequality and Chebyshev inequality

$$\mathbb{E}\left(\|U_j - U_{M,j}\|\right) \leq 2\mathbb{E}^{1/m}\left[\|U_j\|^m\right]\mathbb{P}^{(m-1)/m}\left[\|U_j\| > M\right] \leq 2\mathbb{E}\left[\|U_j\|^m\right]M^{1-m}.$$

Now, to deduce that $R_{1n}(\gamma)$ is uniformly negligible, it suffices to note that under Assumption K-(b), for $m > 7$ and a sufficiently small,

$$M^{1-m} = n^{(1-m)(1/4-a)} = o\left(n^{-1}h^{1/2}p\ln n\right).$$

Clearly, $\sup_\gamma|R_{2n}(\gamma)|$ is of smaller order than $\sup_\gamma|R_{1n}(\gamma)|$.

For the inverse of the variance estimator, for any $\gamma\in\mathcal{S}^p$ let us define

$$\widehat{v}_{N,n}^2(\gamma) = \frac{2}{n(n-1)h}\sum_{j\neq i}\langle U_i,U_j\rangle^2\mathbb{I}_{\{\langle U_i,U_j\rangle^2\leq N\}}K_h^2\left(F_{\gamma,n}(\langle X_i,\gamma\rangle) - F_{\gamma,n}(\langle X_j,\gamma\rangle)\right).$$

Using Hölder inequality, Chebyshev inequality and Cauchy-Schwarz inequality,

$$\begin{aligned}\mathbb{E}\left[\sup_{\gamma}\left|\widehat{v}_n^2(\gamma) - \widehat{v}_{N,n}^2(\gamma)\right| \mid X_1,\cdots,X_n\right] &\leq Ch^{-1}\mathbb{E}\left(\langle U_i,U_j\rangle^2\mathbb{I}_{\{\langle U_i,U_j\rangle^2 > N\}}\right)\\ &\leq h^{-1}\mathbb{E}^{1/s}\left[\langle U_i,U_j\rangle^{2s}\right]\mathbb{P}^{(s-1)/s}\left[\langle U_i,U_j\rangle^{2s} > N^s\right]\\ &\leq h^{-1}\mathbb{E}^2\left[\|U_j\|^{2s}\right]N^{1-s}.\end{aligned}$$

Take $s = 4$, $N = n^{1/4}$ and deduce that the right bound in the last display tends to zero. On the other hand, we apply Hoeffding (1963) inequality for $U-$statistics to control the deviations of $\widehat{v}^2_{N,n}(\gamma) - \mathbb{E}[\widehat{v}^2_{N,n}(\gamma) \mid X_1, \cdots, X_n]$ conditionally on $X_1, \cdots, X_n$. For any fixed γ we have

$$\mathbb{P}\left(n^{1/2}h|\widehat{v}^2_{N,n}(\gamma) - \mathbb{E}[\widehat{v}^2_{N,n}(\gamma) \mid X_1, \cdots, X_n]| \geq t \mid X_1, \cdots, X_n\right) \\ \leq 2\exp\left\{-\frac{[n/2]n^{-1}t^2}{2[\tau^2 + K^2(0)Nn^{-1/2}t/3]}\right\}$$

where τ^2 is the variance of a term in the sum defining $h\widehat{v}^2_{N,n}(\gamma) - \mathbb{E}[h\widehat{v}^2_{N,n}(\gamma) \mid X_1, \cdots, X_n]$. Take $t = n^{1/2-c}h$ for some small $c > 0$ and note that $\tau^2 \leq C$ for some constant independent of γ and h. In the similar way we did for $Q_{M,n}(\gamma)$, applying Lemma 5.4, we obtain an exponential bound for the tail of $\widehat{v}^2_{N,n}(\gamma) - \mathbb{E}[\widehat{v}^2_{N,n}(\gamma) \mid X_1, \cdots, X_n]$ given $X_1, \cdots, X_n$ *uniformly* with respect to γ. This bound is independent of $X_1, \cdots, X_n$. Deduce that

$$\sup_\gamma |\widehat{v}^2_{N,n}(\gamma) - \mathbb{E}[\widehat{v}^2_{N,n}(\gamma) \mid X_1, \cdots, X_n]| = o_{\mathbb{P}}(1),$$

conditionally on $X_1, \cdots, X_n$ and unconditionally. It remains to note that Assumption D-(c) and the first part of Lemma 5.1 guarantee that $\mathbb{E}[\widehat{v}^2_{N,n}(\gamma) \mid X_1, \cdots, X_n]$ stays away from zero. Gathering the results we conclude that $1/\widehat{v}^2_n(\gamma)$ is uniformly bounded in probability. Now the proof is complete. ■

Proof of Theorem 3.3. By Lemma 3.2, if suffices to prove the asymptotic normality of the test statistic T_n defined with $\widehat{\gamma}_n = \gamma_0^{(p)}$. The proof of this asymptotic normality is based on the Central Limit Theorem 5.1 of de Jong (1987). We will apply the result of de Jong conditionally given the values of the covariate sample. Let $x_1, \cdots, x_n$ be an *arbitrary* collection of non-random points in $L^2[0,1]$. Consider $\widetilde{Z}_1, \cdots, \widetilde{Z}_n$ independent random variables with values in $L^2[0,1]$ such that for each i the law of $\widetilde{Z}_i$ is the conditional law of U_i given $X_i = x_i$. Let $F_{\gamma_0^{(p)},n}(\cdot)$ be the empirical d.f. of the sample $\langle x_1, \gamma_0^{(p)}\rangle, \cdots, \langle x_n, \gamma_0^{(p)}\rangle$,

$$K_{h,ij}(\gamma_0^{(p)}) = K_h\left(F_{\gamma_0^{(p)},n}(\langle x_i, \gamma_0^{(p)}\rangle) - F_{\gamma_0^{(p)},n}(\langle x_j, \gamma_0^{(p)}\rangle)\right)$$

and

$$W_{ij} = \frac{1}{n(n-1)}\langle \widetilde{Z}_i, \widetilde{Z}_j\rangle \frac{1}{h}K_{h,ij}(\gamma_0^{(p)}), \quad 1 \leq i \neq j \leq n, \qquad W_{ii} = 0, \quad 1 \leq i \leq n.$$

Hence $Q_n(\gamma_0^{(p)}) = \sum_{i,j} W_{ij}$ and $\widehat{v}^2_n(\gamma_0^{(p)}) = 2n(n-1)h\sum_{i,j} W_{ij}^2$. A crucial remark that is used several times in the following is that the elements of the matrix

$(K_{h,ij}(\gamma_0^{(p)}))$ are the same as those of matrix $(K_h((i-j)/nh)$ up to permutations of lines and columns. Following the notation of de Jong (1987), let

$$\sigma_{ij}^2 = \mathbb{E}(W_{ij}^2) = \mathbb{E}[\langle U_i, U_j\rangle^2 \mid X_i = x_i, X_j = x_j] \frac{K_{h,ij}^2(\gamma_0^{(p)})}{n^2(n-1)^2h^2}$$

and $\sigma(n)^2 = 2\sum_{i\neq j}\sigma_{ij}^2$. Since

$$\mathbb{E}[\langle U_i, U_j\rangle^2 \mid X_1, \cdots, X_n] = \mathbb{E}[\langle U_i, U_j\rangle^2 \mid X_i, X_j] \leq \mathbb{E}[\|U_i\|^2 \mid X_i]\mathbb{E}[\|U_j\|^2 \mid X_j],$$

and $\mathbb{E}[\langle U_i, U_j\rangle^2 \mid X_i, X_j]$ is bounded away from zero by Assumption D-(c), deduce that there exist positive constants $\underline{c}$ and $\overline{c}$ such that

$$\frac{\underline{c}}{n^4h^2}K_{h,ij}^2(\gamma_0^{(p)}) \leq \sigma_{ij}^2 \leq \frac{\overline{c}}{n^4h^2}K_{h,ij}^2(\gamma_0^{(p)}). \tag{5.5.7}$$

Apply Lemma 5.1 with K replaced by K^2 and deduce that for each i,

$$\begin{aligned}\frac{c_1}{n^3h} \leq \frac{\underline{c}}{n^4h^2}\min_{1\leq i\leq n}\sum_{j\neq i}K_h^2((i-j)/nh) &\leq \sum_{1\leq j\leq n, i\neq j}\sigma_{ij}^2 \\ &\leq \frac{\overline{c}}{n^4h^2}\max_{1\leq i\leq n}\sum_{j\neq i}K_h^2((i-j)/nh) \leq \frac{c_2}{n^3h},\end{aligned} \tag{5.5.8}$$

for some constants c_1 and c_2. Moreover, there exist constants $\underline{c}'$ and $\overline{c}'$ such that

$$\underline{c}'n^{-2}h^{-1} \leq \sigma(n)^2 \leq \overline{c}'n^{-2}h^{-1}.$$

It follows that

$$\sigma(n)^{-2}\max_{1\leq i\leq n}\sum_{j=1}^n \sigma_{ij}^2 = O(n^{-1}),$$

and thus Condition 1 in Theorem 5.1 of de Jong (1987) holds true as soon as $\kappa(n) = o(n^{1/2})$. For checking Condition 2 in Theorem 5.1 of de Jong (1987), let us use Hölder inequality with $p = \nu/2$ and $q = \nu/(\nu-2)$, with ν given by Assumption D-(c)-(ii), and Markov inequality to get, for some constant C,

$$\mathbb{E}[\sigma_{ij}^{-2}W_{ij}^2\mathbb{I}_{\{\sigma_{ij}^{-1}|W_{ij}|>\kappa(n)\}}] \leq \mathbb{E}^{2/\nu}[\sigma_{ij}^{-\nu}|W_{ij}|^\nu]\mathbb{P}^{(\nu-2)/\nu}[\sigma_{ij}^{-1}|W_{ij}| > \kappa(n)] \leq C\kappa(n)^{-(\nu-2)/\nu}.$$

That shows that Condition 2 of Theorem 5.1 of de Jong holds true with any $\kappa(n)$ tending to infinity. Finally, let $\mu_1, \cdots, \mu_n$ denote the eigenvalues of the matrix (σ_{ij}). To check Condition 3 of de Jong, use the upper bound of σ_{ij} in (5.5.7) to deduce that there exists a constant C (independent on n and i) such that

$$\sum_{j=1, j\neq i}^n \sigma_{ij} \leq \frac{C}{n^2h}\sum_{j=1,j\neq i}^n K_{h,ij}(\gamma_0^{(p)}).$$

Next, note that if Σ denotes the $n \times n$ matrix with generic element σ_{ij}, following the lines of equation (5.5.5) and using equation (5.5.7), for any $u \in \mathbb{R}^n$,

$$\begin{aligned}
\|\Sigma u\|^2 &\leq \|u\|^2 \left[\max_{1\leq i\leq n} \left(\sum_{j=1, j\neq i}^{n} \sigma_{ij} \right) \right]^2 \\
&\leq c_1 \|u\|^2 \left[\max_{1\leq i\leq n} \left(\sum_{j=1, j\neq i}^{n} \frac{K_h\left((i-j)/nh\right)}{h\, n(n-1)} \right) \right]^2 \\
&\leq c_2 n^{-2} \|u\|^2, \qquad (5.5.9)
\end{aligned}$$

for some constants $c_1, c_2 > 0$. Deduce that

$$\sigma(n)^{-2} \max_{1\leq i\leq n} \mu_i^2 \leq \frac{hn^2}{\underline{c}'} \frac{c_2}{n^2} \to 0,$$

and thus Condition 3 of de Jong (1987) holds true. To complete the proof of the asymptotic normality of the statistic $T_n = nh^{1/2} Q_n(\gamma_0^{(p)})/\widehat{v}_n(\gamma_0^{(p)})$ given the covariate values, note that

$$\sigma^2(n) = \mathbb{E}[Q_n^2(\gamma_0^{(p)}) \mid X_1 = x_1, \cdots, X_n = x_n] = \frac{\mathbb{E}[\widehat{v}_n^2(\gamma_0^{(p)}) \mid X_1 = x_1, \cdots, X_n = x_n]}{n(n-1)h}.$$

Moreover, by direct standard calculations it can be shown that the variance of

$$\frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} \langle \widetilde{Z}_i, \widetilde{Z}_j \rangle^2 \frac{1}{h} K_{h,ij}^2(\gamma_0^{(p)})$$

is of rate $O(h^{-1}n^{-1}) = o(1)$. Deduce that

$$\frac{\widehat{v}_n^2(\gamma_0^{(p)})/n(n-1)h}{\sigma^2(n)} - 1 = o_{\mathbb{P}}(1) \qquad (5.5.10)$$

given $X_1 = x_1, \cdots, X_n = x_n$. The asymptotic normality of T_n given $X_1 = x_1, \cdots, X_n = x_n$ is a consequence of Theorem 5.1 of de Jong and equation (5.5.10). The proof is complete. ■

Proof of Theorem 3.4. The proof is based on inequality (5.3.4). Since $\mathbb{E}(\langle U_1, U_2\rangle^2 \mid X_1, X_2) \geq \underline{\sigma}^2 + r_n^4 \langle \delta(X_1), \delta(X_2)\rangle^2$, clearly the variance estimate $\widehat{v}_n^2(\widetilde{\gamma})$ stays away from zero for all $\widetilde{\gamma}$. On the other hand, by Cauchy-Schwarz and the property of the spectral norm for matrices,

$$\begin{aligned}
\widehat{v}_n^2(\widetilde{\gamma}) &\leq \frac{2n/(n-1)}{n^2 h} \sum_{1\leq i,j\leq n} \|\delta(X_i)\|^2 \|\delta(X_j)\|^2 K_h^2(F_{n,\widetilde{\gamma}}(\langle X_i, \widetilde{\gamma}\rangle) - F_{n,\widetilde{\gamma}}(\langle X_j, \widetilde{\gamma}\rangle)) \\
&\leq |||\mathcal{K}_2|||_2 \sum_{1\leq i\leq n} \|\delta(X_i)\|^4, \qquad (5.5.11)
\end{aligned}$$

where $\mathcal{K}_2$ is the matrix with entries $n^{-2}h^{-1}K_h^2(F_{n,\widetilde{\gamma}}(\langle X_i, \widetilde{\gamma}\rangle) - F_{n,\widetilde{\gamma}}(\langle X_j, \widetilde{\gamma}\rangle))$. By the arguments used in equation (5.5.9), $|||\mathcal{K}_2|||_2 = O_{\mathbb{P}}(n^{-1})$. This together with the finite fourth order moment condition for $\delta(\cdot)$ imply that $\widehat{v}_n^2(\widetilde{\gamma})$ is bounded in probability. Hence it suffices to look at the behavior of $Q_n(\widetilde{\gamma})$. By Lemma 2.1-(B) there exists p_0 and $\widetilde{\gamma} \in B_{p_0} \subset \mathcal{S}^{p_0}$ (p_0 and $\widetilde{\gamma}$ independent of n) such that $\mathbb{E}[\delta(X) \mid \langle X, \widetilde{\gamma}\rangle] \neq 0$. Hereafter, $\widetilde{\gamma}$ is supposed to have this property. Let $V_{ni} = F_{n,\widetilde{\gamma}}(\langle X_i, \widetilde{\gamma}\rangle)$. We can write

$$\begin{aligned}
Q_n(\widetilde{\gamma}) &= \frac{1}{n(n-1)h}\sum_{i\neq j}\langle U_i^0, U_j^0\rangle K_h(V_{ni} - V_{nj}\rangle) \\
&\quad + \frac{2r_n}{n(n-1)h}\sum_{i\neq j}\langle U_i^0, \delta(X_j)\rangle K_h(V_{ni} - V_{nj}) \\
&\quad + \frac{r_n^2}{n(n-1)h}\sum_{i\neq j}\langle \delta(X_i), \delta(X_j)\rangle K_h(V_{ni} - V_{nj}) \\
&=: Q_{0n}(\widetilde{\gamma}) + 2r_n Q_{1n}(\widetilde{\gamma}) + r_n^2 Q_{2n}(\widetilde{\gamma}).
\end{aligned}$$

Since $\widetilde{\gamma}$ is fixed, $Q_{0n}(\widetilde{\gamma}) = O_{\mathbb{P}}(n^{-1}h^{-1/2})$ (cf. proof of Theorem 3.3). Next, let us follow Guerre and Lavergne (2005), denote by $\mathbb{E}_n$ the conditional expectation given $X_1, \cdots, X_n$ and define

$$\overline{\delta_n}(X_i) = \frac{1}{n(n-1)h}\sum_{j=1,\, j\neq i}^{n}\delta(X_j)K_h(V_{ni} - V_{nj}), \qquad \overline{\delta} = (\delta(X_1), \cdots, \delta(X_n))'.$$

Then Marcinkiewicz-Zygmund inequality, see Chow and Teicher (1988, Theorem 2, p. 386), and Cauchy-Schwarz and Jensen inequalities imply

$$\begin{aligned}
\mathbb{E}_n\left|\sum_{i=1}^{n}\langle U_i^0, \overline{\delta_n}(X_i)\rangle\right| &\leq c\,\mathbb{E}_n\left|\sum_{i=1}^{n}\left|\langle U_i^0, \overline{\delta_n}(X_i)\rangle\right|^2\right|^{1/2} \leq c\,\mathbb{E}_n\left|\sum_{i=1}^{n}\|U_i^0\|^2\|\overline{\delta_n}(X_i)\|^2\right|^{1/2} \\
&\leq c\left\{\sum_{i=1}^{n}\mathbb{E}_n\left(\|U_i^0\|^2\right)\|\overline{\delta_n}(X_i)\|^2\right\}^{1/2} \leq c\,C_2^{1/\nu}\left\{\sum_{i=1}^{n}\|\overline{\delta_n}(X_i)\|^2\right\}^{1/2} \\
&= c\,C_2^{1/\nu}\|\mathcal{K}_3\overline{\delta}\| \leq c\,C_2^{1/\nu}n^{1/2}|||\mathcal{K}_3|||_2\left\{\frac{1}{n}\sum_{i=1}^{n}\|\delta(X_i)\|^2\right\}^{1/2},
\end{aligned}$$

for $\mathcal{K}_3$ a matrix with the same elements as the matrix $\mathcal{K}$ defined in equation (5.5.4) up to permutations of lines and columns, and C_2 and ν the constants in Assumption D, and c some constant independent of n. Since $|||\mathcal{K}|||_2 = |||\mathcal{K}_3|||_2 = O_{\mathbb{P}}(n^{-1})$, deduce that $Q_{1n}(\widetilde{\gamma}) = O_{\mathbb{P}}(n^{-1/2})$ conditionally on $X_1, \cdots, X_n$. Now, let us investigate $Q_{2n}(\widetilde{\gamma})$. By an inequality like in equation (5.5.11) and the moment conditions on $\delta(\cdot)$ it is easy to bound $Q_{2n}(\widetilde{\gamma})$ in probability. It remains to show that it is bounded away from zero. Let $V_i = F_{\widetilde{\gamma}}(\langle X_i, \widetilde{\gamma}\rangle)$, so that $V_1, \cdots, V_n$ are independent uniform variables on $[0,1]$, and

$$Q'_{2n}(\widetilde{\gamma}) = \frac{1}{n^2 h} \sum_{1\le i,j\le n} \langle \delta(X_i), \delta(X_j)\rangle K_h(V_{ni} - V_{nj}),$$

$$Q''_{2n}(\widetilde{\gamma}) = \frac{1}{n^2 h} \sum_{1\le i,j\le n} \langle \delta(X_i), \delta(X_j)\rangle K_h(V_i - V_j),$$

$$Q^{\star}_{2n}(\widetilde{\gamma}) = \frac{1}{n(n-1)h} \sum_{1\le i\neq j\le n} \langle \delta(X_i), \delta(X_j)\rangle K_h(V_i - V_j).$$

We have

$$Q'_{2n}(\widetilde{\gamma}) - \frac{n-1}{n} Q_{2n}(\widetilde{\gamma}) = Q''_{2n}(\widetilde{\gamma}) - \frac{n-1}{n} Q^{\star}_{2n}(\widetilde{\gamma}) = \frac{K(0)}{n^2 h}\sum_{i=1}^{n} \|\delta(X_i)\|^2 = O_{\mathbb{P}}(n^{-1}h^{-1}) = o_{\mathbb{P}}(1).$$

Next we show that $Q'_{2n}(\widetilde{\gamma}) - Q''_{2n}(\widetilde{\gamma}) = o_{\mathbb{P}}(1)$. If K satisfies a Lipschitz condition and $nh^4 \to \infty$, by Cauchy-Schwarz inequality, for some constant $C > 0$

$$|Q'_{2n}(\widetilde{\gamma}) - Q''_{2n}(\widetilde{\gamma})| \;\le\; \frac{C\Delta_n}{h^2}\left[\frac{1}{n}\sum_{1\le i\le n} \|\delta(X_i)\|\right]^2 = o_{\mathbb{P}}(1),$$

where $\Delta_n = \sup_{1\le i\le n} |V_{ni} - V_i|$. Note that $\Delta_n \le \sup_{t\in\mathbb{R}} |F_{n,\widetilde{\gamma}}(t) - F_{\widetilde{\gamma}}(t)| = O_{\mathbb{P}}(n^{-1/2})$. Conclude that $Q_{2n}(\widetilde{\gamma}) - Q^{*}_{2n}(\widetilde{\gamma}) = o_{\mathbb{P}}(1)$, so that is suffices to investigate $Q^{*}_{2n}(\widetilde{\gamma})$. It is easy to show that the variance of $Q^{*}_{2n}(\widetilde{\gamma})$ tends to zero, so that it remains to show that the expectation of $Q^{*}_{2n}(\widetilde{\gamma})$ stay away from zero. Let $\overline{\delta}(t,v) = \mathbb{E}[\delta(X_j)(t) \mid V_j = v]$ and note that $0 < \iint_{[0,1]\times[0,1]} |\overline{\delta}(t,v)|^2 dvdt < \infty$. By the Inverse Fourier Transform formula and repeated applications of Fubini's theorem we get

$$\begin{aligned}
\mathbb{E}[Q^{*}_{2n}(\widetilde{\gamma})] &= \mathbb{E}\left[\langle \delta(X_i), \delta(X_j)\rangle h^{-1} K_h(V_i - V_j)\right] \\
&= \mathbb{E}(\langle \delta(X_i), \mathbb{E}[\delta(X_j) h^{-1} K_h(V_i - V_j) \mid X_i]\rangle) \\
&= \int_{[0,1]} \mathbb{E}\left(\delta(X)(t) \int_{\mathbb{R}} \exp\{2\pi i s V\}\mathcal{F}[\overline{\delta}(t,\cdot)](-s)\mathcal{F}[K](hs)ds\right) dt \\
&= \int_{[0,1]} \left[\int_{\mathbb{R}} \|\mathcal{F}[\overline{\delta}(t,\cdot)](s)\|^2 \mathcal{F}[K](hs)ds\right] dt.
\end{aligned}$$

When $h \to 0$, by Lebesgue dominated convergence theorem and Plancherel theorem applied for the integral inside the square brackets,

$$\mathbb{E}[Q^{*}_{2n}(\widetilde{\gamma})] \to \int_{[0,1]}\int_{[0,1]} |\overline{\delta}(t,v)|^2 dvdt.$$

Deduce that $\mathbb{P}[c^{-1} \leq Q_{2n}(\widetilde{\gamma}) \leq c] \to 1$ for some constant $c > 0$. Gathering the results conclude that for any $C > 0$, $\mathbb{P}[T_n \geq C] \to 1$. ■

Proof of Corollary 3.5. a) Let $\widehat{x}_{ik} = \int_{[0,1]} X_i(t)\widehat{\rho}_k(t)dt$, so that $\langle X_i, \gamma\rangle_n = \sum_{k=1}^p \widehat{x}_{ik}\gamma_k$. Note that $\widehat{x}_{ik}$, $1 \leq k \leq p$, $1 \leq i \leq n$ are measurable functions of $X_1, \cdots, X_n$. Now, let $\widehat{F}_{\gamma,n}$ denote the empirical distribution function of the sample $\langle X_1, \gamma\rangle_n, \cdots, \langle X_n, \gamma\rangle_n$. Note that the elements of the matrices $(K_h(\widehat{F}_{\gamma,n}(\langle X_i, \gamma\rangle_n) - \widehat{F}_{\gamma,n}(\langle X_j, \gamma\rangle_n)))$ and $(K((i-j)/nh))$ are the same up to permutations of lines and columns. Given that in the proofs of Lemma 3.1 and Theorem 3.3 the arguments were provided conditionally on $X_1, \cdots, X_n$, it is quite clear that the conclusion of Theorem 3.3 remains true if the $\langle X_i, \gamma\rangle$'s are everywhere replaced by the $\langle X_i, \gamma\rangle_n$.

b) Similarly, all but one of the arguments in the proof of Theorem 3.4 applies with the $\langle X_i, \gamma\rangle_n$'s. It only remains to investigate the counterpart of $Q_{2n}(\widetilde{\gamma})$ that was the leading term in $Q_n(\widetilde{\gamma})$. For this purpose, note that for any γ, $\langle X_i, \gamma\rangle_n = \langle X_i, \gamma\rangle + \langle X_i, \Delta_{n,\gamma}\rangle$ where

$$\Delta_{n,\gamma}(t) = \sum_{k=1}^p \gamma_k[\widehat{\rho}_k(t) - \rho_k(t)], \qquad t \in [0,1].$$

For an integral operator $(\Psi v)(t) = \int \psi(t,s)v(s)ds$ with $\int\int \psi^2(t,s)dtds < \infty$, consider the operator norm $\|\Psi\|_S$ defined by $\|\Psi\|_S^2 = \int\int \psi^2(t,s)dtds$. Under Assumption D-(a) and the moment assumption on $\|X\|$,

$$\|\widehat{\Gamma} - \Gamma\|_S = O_{\mathbb{P}}(1/\sqrt{n}),$$

see for instance Bosq (2000) or Horváth and Kokoszka (2012). Next, by Cauchy-Schwarz inequality, Lemma 4.3 in Bosq (2000) or Lemma 2.3 in Horváth and Kokoszka (2012), and the fact that the spectral norm of the operator $\widehat{\Gamma} - \Gamma$ is smaller or equal to $\|\widehat{\Gamma} - \Gamma\|_S$,

$$\int_{[0,1]} \Delta_{n,\gamma}^2(t)dt \leq \left[\sum_{k=1}^p \gamma_k^2\right] \sum_{k=1}^p \|\widehat{\rho}_k - \rho_k\|^2 \leq p\frac{8^{1/2}}{\varsigma_p}\|\widehat{\Gamma} - \Gamma\|_S^2,$$

where $\varsigma_p = \min_{1\leq j\leq p}(\lambda_j - \lambda_{j+1})$. Then the lower bound for the spacing between the eigenvalues implies

$$\sup_{\gamma\in\mathcal{S}^p} \int_{[0,1]} \Delta_{n,\gamma}^2(t)dt \leq cp^{\eta+1}\|\widehat{\Gamma} - \Gamma\|_S^2,$$

for some constant $c > 0$. Deduce that

$$\sup_{\gamma\in\mathcal{S}^p} \max_{1\leq i\leq n} |\langle X_i, \gamma\rangle_n - \langle X_i, \gamma\rangle| \leq \max_{1\leq i\leq n} \|X_i\| c^{1/2} p^{(\eta+1)/2} \|\widehat{\Gamma} - \Gamma\|_S = O_{\mathbb{P}}(p^{(\eta+1)/2}\ln n/\sqrt{n}),$$

where for the last equality we used the condition $\mathbb{E}[\exp(\varrho\|X\|)] < \infty$ to deduce that $\max_{1\leq i\leq n}\|X_i\| = O_{\mathbb{P}}(\ln n)$. Let $b_n \downarrow 0$ such that $b_n\sqrt{n}/[p^{(\eta+1)/2}\ln n] \to \infty$ and define the event

$$\mathcal{E}_n = \{\sup_{\gamma\in\mathcal{S}^p}\max_{1\leq i\leq n}|\langle X_i,\gamma\rangle_n - \langle X_i,\gamma\rangle| \leq b_n\}$$

so that $\mathbb{P}(\mathcal{E}_n^c) \to 0$. On the set $\mathcal{E}_n$, for any $\gamma \in \mathcal{S}^p$ and $t \in \mathbb{R}$ we can write

$$\widehat{F}_{\gamma,n}(t) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{\{\langle X_i,\gamma\rangle_n\leq t\}} \leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{\{\langle X_i,\gamma\rangle\leq t+b_n\}} = F_{\gamma,n}(t+b_n),$$

and similarly, $\widehat{F}_{\gamma,n}(t) \geq F_{\gamma,n}(t-b_n)$. Deduce that on $\mathcal{E}_n$,

$$\begin{aligned}&\left|\widehat{F}_{\widetilde{\gamma},n}(\langle X_i,\widetilde{\gamma}\rangle_n) - \widehat{F}_{\widetilde{\gamma},n}(\langle X_j,\widetilde{\gamma}\rangle_n)\right|\\ &\leq \max\{|F_{\widetilde{\gamma},n}(\langle X_i,\widetilde{\gamma}\rangle+b_n)-F_{\widetilde{\gamma},n}(\langle X_j,\widetilde{\gamma}\rangle-b_n)|\,,\ \ |F_{\widetilde{\gamma},n}(\langle X_i,\widetilde{\gamma}\rangle-b_n)-F_{\widetilde{\gamma},n}(\langle X_j,\widetilde{\gamma}\rangle+b_n)|\}.\end{aligned}$$

On the other hand,

$$\begin{aligned}|F_{\widetilde{\gamma},n}(\langle X_i,\widetilde{\gamma}\rangle+b_n)-F_{\widetilde{\gamma},n}(\langle X_j,\widetilde{\gamma}\rangle-b_n)| &\leq |F_{\widetilde{\gamma},n}(\langle X_i,\widetilde{\gamma}\rangle+b_n) - F_{\widetilde{\gamma}}(\langle X_i,\widetilde{\gamma}\rangle+b_n)|\\ +\,|F_{\widetilde{\gamma}}(\langle X_i,\widetilde{\gamma}\rangle+b_n) - F_{\widetilde{\gamma}}(\langle X_i,\widetilde{\gamma}\rangle-b_n)| &+ |F_{\widetilde{\gamma},n}(\langle X_i,\widetilde{\gamma}\rangle-b_n) - F_{\widetilde{\gamma}}(\langle X_i,\widetilde{\gamma}\rangle-b_n)|\\ &\leq 2\sup_{t\in\mathbb{R}}|F_{\widetilde{\gamma},n}(t) - F_{\widetilde{\gamma}}(t)| + 2b_n\sup_{t\in\mathbb{R}}f_{\widetilde{\gamma}}(t)\\ &= O_{\mathbb{P}}(n^{-1/2}+b_n) = O_{\mathbb{P}}(b_n).\end{aligned}$$

From this and the Lipschitz condition on K, deduce that

$$\left|K_h(\widehat{F}_{\widetilde{\gamma},n}(\langle X_i,\widetilde{\gamma}\rangle_n)-\widehat{F}_{\widetilde{\gamma},n}(\langle X_j,\widetilde{\gamma}\rangle_n)) - K_h(F_{\widetilde{\gamma},n}(\langle X_i,\widetilde{\gamma}\rangle)-F_{\widetilde{\gamma},n}(\langle X_j,\widetilde{\gamma}\rangle))\right| = O_{\mathbb{P}}(b_nh^{-1}).$$

Let $\widehat{Q}_{2n}(\widetilde{\gamma})$ be defined like $Q_{2n}(\widetilde{\gamma})$ but with $\langle X_i,\widetilde{\gamma}\rangle$'s replaced by $\langle X_i,\widetilde{\gamma}\rangle_n$'s. Deduce from above

$$\left|\widehat{Q}_{2n}(\widetilde{\gamma}) - Q_{2n}(\widetilde{\gamma})\right| \leq O_{\mathbb{P}}(b_nh^{-2})\left[\frac{1}{n}\sum_{1\leq i\leq n}\|\delta(X_i)\|\right]^2 = o_{\mathbb{P}}(1),$$

provided $b_n\sqrt{n}/[p^{(\eta+1)/2}\ln n] \to \infty$ and $b_n = o(h^2)$. The conclusion follows. ■

Proof of Theorem 3.6. The idea is to show the asymptotic normality of T_n^b conditionally on the observed data adapting the steps used to derive the asymptotic normality for T_n. Consider the event $\mathcal{A}_n = \{\max_{1\leq i\leq n} \|U_i\| \leq M\}$ with $M = n^{1/4-a}$ for some small a. Assumption D-(a) guarantees $\mathbb{P}(\mathcal{A}_n^c) \to 0$. Define

$$h_{i,j}^b = \frac{\zeta_i\, \zeta_j}{n(n-1)h} C_{n,ij},$$

where

$$C_{n,ij} = \frac{\langle U_i \mathbb{I}_{\{\|U_i\|\leq M\}}, U_j \mathbb{I}_{\{\|U_j\|\leq M\}}\rangle}{M^2} K_h\left(F_{\gamma,n}(\langle x_i,\gamma\rangle) - F_{\gamma,n}(\langle x_j,\gamma\rangle)\right).$$

Let $Q_n^b(\gamma)$ be the bootstrap version of $Q_n(\gamma)$, and let

$$Q_{M,n}^b(\gamma) = \frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} h_{i,j}^b, \quad \gamma \in \mathcal{S}^p.$$

Note that for any $t > 0$

$$\mathbb{P}\left[\sup_\gamma \left|Q_n^b(\gamma) - Q_{M,n}^b(\gamma)\right| > t \mid U_1, X_1, \cdots, U_n, X_n\right] \leq \mathbb{P}(\mathcal{A}_n^c) \to 0. \qquad (5.5.12)$$

Define the quantities A_n^b, B_n^b, C_n^b and D_n^b like in (5.5.1)-(5.5.2) with $h_{i,j}$ replaced by $h_{i,j}^b$ and the expectations replaced by the conditional expectations given $(U_1, X_1), \cdots, (U_n, X_n)$. It is easy to check that the same upper bounds like in Lemma 5.3 could be derived on the event $\mathcal{A}_n$. Then equation (5.5.12) and the exponential inequality from Lemma 5.2 applied like in Lemma 3.1 yields, for any $C > 0$,

$$\mathbb{P}\left[\sup_\gamma \left|Q_n^b(\gamma)\right| > Cp\ln n/nh^{1/2} \mid U_1, X_1, \cdots, U_n, X_n\right] \to 0 \quad \text{in probability}.$$

Deduce that $\mathbb{P}(\widehat{\gamma}_n \neq \gamma_0^{(p)} \mid U_1, X_1, \cdots, U_n, X_n) \to 0$, in probability. The second part of lemma 3.1 follows from similar arguments. It remains to derive the reconsider the steps of Theorem 3.3 with $\widetilde{Z}_i$ replaced by ζ_i and $K_{h,ij}(\gamma_0^{(p)})$ by $K_{h,ij}^b(\gamma_0^{(p)}) = \langle U_i, U_j\rangle K_{h,ij}(\gamma_0^{(p)})$. Consequently W_{ij} becomes $W_{ij}^b = n^{-1}(n-1)^{-1}h^{-1}\zeta_i\zeta_j K_{h,ij}^b(\gamma_0^{(p)})$, σ_{ij} is replaced by $(\sigma_{ij}^b)^2 = [K_{h,ij}^b(\gamma_0^{(p)})]^2/[n^2(n-1)^2h^2]$ and $\sigma(n)^2$ is now $\sigma^b(n)^2 = 2\sum_{i\neq j}(\sigma_{ij}^b)^2$. Define the set $\mathcal{E}_{1n} = \{\sigma^b(n)^2 \geq \underline{\sigma}^2\}$ where $\underline{\sigma}^2$ is the lower bound in Assumption D-(c)-(i). Since $\lim_n \mathbb{E}(\sigma^b(n)^2) \geq 2\underline{\sigma}^2$ and the variance of $\sigma^b(n)^2$ tends to zero, deduce that $\mathbb{P}(\mathcal{E}_{1n}^c) \to 0$. Next, define $\mathcal{E}_{2n} = \{\sigma^b(n) \leq 2C_2^{2/\nu}\}$, with C and ν defined in Assumption D-(c)-(ii) and note that $\lim_n \mathbb{E}(\sigma^b(n)^2) \leq 2C_2^{4/\nu}$ so that $\mathbb{P}(\mathcal{E}_{2n}^c) \to 0$. On $\mathcal{E}_{1n} \cap \mathcal{E}_{2n}$, Conditions 1 and 2 of Theorem 5.1 of de Jong (1987) are clearly satisfied. For checking Condition 3, let

$\mathcal{K}^b$ denote the matrix with generic element $\mathcal{K}^b_{ij} = K^b_{h,ij}(\gamma_0^{(p)})$ if $i \neq j$ and $\mathcal{K}^b_{ij} = 0$ otherwise. Recall that $\mathbb{E}_n$ stands for the conditional expectation given $X_1, \cdots, X_n$ and note that $\mathbb{E}_n(\|U_i\|\|U_j\|) \leq \mathbb{E}^{1/2}(\|U_i\|^2 \mid X_i)\mathbb{E}^{1/2}(\|U_j\|^2 \mid X_j) \leq C_2^{4/\nu}$. Using the conditional independence between any U_i and the rest of the sample, for any $w \in \mathbb{R}^n$ with $\|w\| = 1$,

$$
\begin{aligned}
\mathbb{E}_n \left\|\mathcal{K}^b w\right\|^2 &\leq \sum_{i=1}^n \mathbb{E}(\|U_i\|^2 \mid X_i)\mathbb{E}_n \left(\sum_{j=1, j\neq i}^n \|U_j\| K^b_{h,ij}(\gamma_0^{(p)}) w_j \right)^2 \\
&\leq \frac{C_2^{6/\nu}}{h^2 n^2 (n-1)^2} \sum_{i,j,k=1}^n K_h\left((i-j)/nh\right) K_h\left((i-k)/nh\right) |w_j w_k| \\
&\leq C_2^{6/\nu} K^2(0) \frac{1}{h^2 n(n-1)^2} \sum_{j,k=1}^n |w_j w_k| \\
&\leq \frac{C_3}{n^2}\frac{1}{nh^2}, \qquad \text{[Cauchy-Schwarz inequality]}
\end{aligned}
$$

where $C_3 > 0$ is some constant. Deduce that $\mathbb{E}\left\|\mathcal{K}^b w\right\|^2 = o(n^{-2})$. Next, define $\mathcal{E}_{3n} = \{|||\mathcal{K}^b|||_2 \leq 1/n\}$, and deduce from above that $\mathbb{P}(\mathcal{E}_{3n}^c) \to 0$. Condition 3 in Theorem 5.1 of de Jong (1987) is clearly satisfied on $\mathcal{E}_{3n}$ and hence that CLT could be applied on the event $\mathcal{E}_n = \mathcal{E}_{1n} \cap \mathcal{E}_{2n} \cap \mathcal{E}_{3n}$ which has a probability tending to one. Finally, it remains to note that equation (5.5.10) holds on $\mathcal{E}_n$. The arguments for the test statistic built with the estimated FPC basis (that is not changed in the bootstrap procedure) are similar. ■

Chapitre 6

Conclusion

Pour conclure, je présente les perspectives de travail. Les données fonctionnelles sont le fil conducteur de cette thèse. Je me suis en premier lieu intéressé au modèle fonctionnel d'équations estimantes, j'avais en tête de construire un test d'adéquation pour ce modèle. Je me suis alors rendu compte qu'il existait très peu de tests en données fonctionnelles, c'est pourquoi j'ai commencé par le modèle linéaire fonctionnel. J'ai alors décrit différents tests en données fonctionnelles. Le test d'adéquation à des modèles paramétriques a été présenté dans le cas du modèle linéaire fonctionnel, ce test s'adapte aussi aux modèles linéaires fonctionnelles généralisés. Il est intéressant d'essayer de pouvoir l'adapter à des modèles plus généraux tel le modèle fonctionnel d'équations estimantes présenté dans le chapitre 2. Une autre perspective serait d'appliquer la méthode du lissage par les plus proches voisins dans les tests d'adéquation de différents modèles paramétriques et de comparer les résultats avec ceux existants.
Les critères de sélection théorique des différents paramètres (paramètre fenêtre, paramètre de troncature) est un sujet vaste et complexe qui me semble être une bonne perspective de travail. En effet, peu de travail a été effectué sur ce domaine. Enfin, j'aimerai m'intéresser à une représentation non-linéaire des données fonctionnelles, ce qui n'est pas le cas avec l'analyse en composantes principales fonctionnelles. Le travail de Chen et Müller (2012) sur les variétés est à mon avis un bon outil de départ.

Table des figures

Liste des tableaux

Chapitre 7

Références

AGUILERA, A., OCANA, F. AND VALDERRAMA, M. (2008) Estimation of functional regression models for functional responses by wavelet approximation. *International Workshop on Functional and Operatorial Statistics 2008 Proceedings, Functional and operatorial statistics, Dabo-Niang and Ferraty (Eds.), Physica-Verlag, Springer.*

ANTOCH, J., PRCHAL, L., DE ROSA, M. AND SARDA, P. (2008) Functional linear regression with functional response : application to prediction of electricity consumption. *International Workshop on Functional and Operatorial Statistics 2008 Proceedings, Functional and operatorial statistics, Dabo-Niang and Ferraty (Eds.), Physica-Verlag, Springer.*

ARCONES, M., AND GINÉ, E. (1993). Limit Theorems for $U-$processes *Ann. Probab.* **22(3)**, 1494–1542.

BIERENS, H.J. (1990). A consistent conditional moment test of functional form. *Econometrica* **58**, 1443–1458.

BOSQ, D. (2000). *Linear Processes in Function Spaces : Theorey and Applications*. Lecture Notes in Statistics (v. 149). Springer-Verlag, New-York.

BROWN, B.M. (1971). Martingale central limit theorems. *Ann. Math. Statist.* **42**, 59–66.

CAI, T., AND HALL, P. (2006). Prediction in functional linear regression. *Annals of Statistics* **34**, 2159–2179.

CARDOT, H., FERRATY, F., MAS, A., AND SARDA, P. (2003). Testing Hypotheses in the Functional Linear Model. *Scandinavian Journal of Statistics* **30**, 241–255.

CARDOT, H., FERRATY, F., AND SARDA, P. (1999). Functional Linear Model. *Statis. Probab. Lett.* **45**, 11–22.

Cardot, H., Goia, P., and Sarda, P. (2004). Testing for no effect in functional linear regression models, some computational approaches. *Communications in Statistics - Simulation and Computation* **33**, 179–199.

Cardot, H., and Sarda, P. (2010). Functional linear regression. *Handbook on Functional Data Analysis and related fields* Oxford University Press.

Chen, D., Hall, P., Müller, H-G. (2011). Single and multiple index functional regression models with nonparametric link. *Annals of Statistics* **39**, 1720–1747.

Chen, D., and Müller, H.G. (2012). Nonlinear manifold representations for functional data. *Annals of Statistics* **40**, 1–29.

Chiou, J-M., Müller, H-G. (2007). Diagnostics for Functional Regression via Residual Processes. *Comput. Statist. Data Anal.* **15**, 4849–4863.

Chiou, J-M., Müller, H-G. and Wang J-L. (2004), Functional response models. *Statistica Sinica* **14**, 659–677.

Chow, Y.S., and Teicher, H. (1997). *Probability Theory : Independence, Interchangeability, Martingales* (3rd ed.) Springer-Verlag, New-York.

Cover, T.M. (1967). The number of linearly inducible orderings in $d-$space. *SIAM J. Appl. Math.* **15**, 434–439.

Crambes, C., Kneip, A., and Sarda, P. (2008). Smoothing splines estimators for functional linear regression. *Annals of Statistics* **37**, 35–72.

Crambes, C., and Mas, A. (2011). Asymptotics of prediction in the functional linear regression with functional outputs. To appear in Bernoulli (2012) arXiv :0910.3070v3 [math.ST]

Cuesta-Albertos, J.A., del Barrio, E., Fraiman, R. and Matran, C. (2007). . *Computational statistics and data analysis* **51**, 4864–4877.

Cuesta-Albertos, E., Fraiman, R., and Ransford, T. (2007). A sharp form of the Cramér-Wold theorem. *J. Theoret. Probab.* **20**, 201–209.

Cuevas, A., Febrero, M., Fraiman, R. (2002). Linear functional regression : The case of fixed design and functional response. *Canadian Journal of Statistics* **30**, 285–300.

Delsol, L., Ferraty, F., and Vieu, P. (2011). Structural test in regression on functional variables. *Journal of Multivariate Analysis* **102**, 422–447.

de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Related Fields* **25**, 261–277.

de la Pẽna, V., and Giné, E. (1999). *Decoupling : from dependence to independence.* Lecture Notes in Mathematics 1617. Springer-Verlag, New-York.

Dominguez, M.A. and Lobato, I.N. (2004). Consistent estimation of models defined by conditional moment restrictions. *Econometrica* **72**, 1601–1615.

Dou, W.W., Pollard, D., and Zhou, H.H. (2012). Estimation in functional regression for general exponential families. *Journal of Multivariate Analysis* **40**, 2421–2451.

Einmahl, U., and Mason, D.M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Annals of Statistics* **33**, 1380–1403.

Escanciano, J.C. (2006). A Consistent diagnostic test for regression models using projections. *Econometric Theory* **22**, 1030–1051.

Fan, Y., and Li, Q. (1996). Consistent model specification tests : omitted variables and semiparametric functional forms. *Econometrica* **64**, 865–890.

Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *J. Stat. Planning and Inference* **140**, 335–352.

Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2011). Kernel regression with functional response. *Electron. J. Stat.* **5**, 159–171.

Ferraty, F., Van Keilegom, I., and Vieu, P. (2012). Regression when both response and predictor are functions. *J. Multivariate Anal.* **109**, 10–28.

Ferraty, F. (Ed.) (2011). *Recent Advances in Functional Data Analysis and Related Topics.* Springer-Verlag Berlin Heidelberg.

Ferraty, F., and Vieu, P. (2006). *Nonparametric Functional Data Analysis : Theory and Practice*. Springer, Berlin.

Gabrys,R., Horváth, L., and Kokoszka, P. (2010). Tests for error correlation in the functional linear model. *Journal of the American Statistical Association* **105**, 1113-1125.

Gao, J., and Gijbels, I. (2008). Bandwidth selection in nonparametric kernel testing. *J. Amer. Statist. Assoc.* **103**, 1584–1594.

García-Portugués, E., González-Manteiga, W., and Febrero-Bande, M. (2012) A goodness-of-fit test for the functional linear model with scalar response. arXiv :1205.6167v3 [stat.ME]

Giné, E. and Guillou, A. (2001). On consistency of kernel density estimators for randomly censored data : rates holding uniformly over adaptive intervals. *Ann. I. H. Poincaré* **37**, 503–522.

Giné, E., Latała, R., and Zinn, J. (2000). Exponential and moment inequalities for $U-$statistics. In *High Dimensional II* 13–38. Progr. Probab. **47**. Birkhäuser, Boston.

Gozalo, P.L. (1997). Nonparametric bootstrap analysis with applications to demographic effects in demand functions. *Journal of Econometrics* **81**, 357-393.

Guerre, E., and Lavergne, P. (2002). Optimal minimax rates for nonparametric specification testing in regression models. *Econometric Theory* **18**, 1139–1171.

Guerre, E., and Lavergne, P. (2005). Data-driven rate-optimal specification testing in regression models. *Annals of Statistics* **33**, 840–870.

HALL, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Amer. Statist. Assoc.* **14**, 1–16.

HALL, P., AND HOROWITZ, J.L. (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics* **35**, 70–91.

HALL, P., AND HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *J. Roy. Stat. Soc. Ser. B* **68**, 109–126.

HALL, P., AND HOSSEINI-NASAB, M. (2009). Theory for high-order bounds in functional principal components analysis. *Math. Proc. Cambridge Philos. Soc.* **146**, 225–256.

HANSEN, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50**, 1029–1054.

HÄRDLE, W., AND MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics* **21**, 1296–1947.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13–30.

HOROWITZ, J.L., AND SPOKOINY, V.G. (2001). An adaptive, rate-optimal test of a parametric model against a nonparametric alternative. *Econometrica* **69**, 599–631.

HORVÁTH, L., AND KOKOSZKA, P. (2012). *Inference for Functional Data with Applications.* Springer Series in Statistics. Springer, New-York.

HORVÁTH, L., AND REEDER, R. (2011). A test of significance in functional quadratic regression. arXiv :1105.0014v1 [math.ST], *forthcoming in Bernoulli Journal*

HORVÁTH, L., AND REEDER, R. (2011). Detecting changes in functional linear models. arXiv :1105.0015v2 [math.ST]

KITAMURA, Y., TRIPATHI, G. AND AHN, H. (2004). Empirical likelihood-based inference in conditional moment restriction models. *Econometrica* **72**, 1667–1714.

KOKOSZKA, P. AND AL. (2008). Testing for lack of dependence in the functional linear model. *Canadian Journal of Statistics* **36**, 207–222.

KOSOROK, M.R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer Series in Statistics. Springer-Verlag, New-York.

LAVERGNE, P. AND PATILEA, V. (2008). Breaking the curse of dimensionality in nonparametric testing. *Journal of Econometrics* **143**, 103–122.

LEDOUX, M. AND TALAGRAND, M. (1991). *Probability in Banach spaces. Isoperimetry and processes.* Springer-Verlag, Berlin.

LI, Q., AND WANG, S. (1998). A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics* **87**, 145–165.

LIAN, H. (2011). Convergence of functional $k-$nearest neighbor regression estimate with functional responses. *Electron. J. Statist.* **5**, 31–40.

MAJOR, P. (2006). An estimate on the supremum of a nice class of stochastic integrals and U-statistics. *Probability Theory and Related Fields* **134**, 489–537.

MAMMEN, E. (1993). Bootstrap and Wild Bootstrap for High Dimensional Linear Models. *Annals of Statistics* **21**, 255–285.

MÜLLER, H-G. AND STADTMÜLLER, U. (2005). Generalized functional linear models. *Annals of Statistics* **33**, 774–805.

NEWEY, W.K. (1993). *Efficient estimation of models with conditional moment restrictions.* Handbook of Statistics, Vol 11 : Econometrics. Madala, Rao and Vinod, eds, Amsterdam : North-Holland.

NEWEY, W.K. AND MCFADDEN, D. (1994). *Large sample estimation and hypothesis testing* Handbook of Econometrics, Vol 4. R. Engle and D. McFadden, eds, Amsterdam : North-Holland.

NOLAN, D., AND POLLARD, D. (1987). $U-$processes : Rates of convergence. *Annals of Statistics* **15**, 780–799.

PARTHASARATHY, K.R. (1967). *Probability measures on metric spaces.* A.M.S. New-York.

PATILEA, V., SÁNCHEZ-SELLERO, C., AND SAUMARD, M. (2012). Projection-based nonparametric goodness-of-fit testing with functional covariates. arXiv :1205.5578 [math.ST]

RAMSAY, J., AND SILVERMAN, B.W. (2005). *Functional Data Analysis* (2nd ed.). Springer-Verlag, New York.

REISS, P.T. AND OGDEN, R.T. (2007). Functional principal component regression and functional partial least squares. *J. Amer. Statist. Assoc.* **102**, 984–996.

RUDIN, W. (1987). *Real and complex analysis.* McGraw-Hill.

SHORACK, G., AND WELLNER, J.A. (1986). *Empirical Processes with Applications in Statistics.* John Wiley & Sons, Inc.

SMITH, R.J. (2007). Efficient information theoretic inference for conditional moment restrictions. *Journal of Econometrics* **138**, 430–460.

STUTE, W. (1984). Asymptotic normality of nearest neighbor regression function estimates. *Ann. Statist.* **12**, 917–926.

STUTE, W. (1997). Nonparametric models checks for regression. *Annals of Statistics* **25**, 613–641.

STUTE, W., GONZÁLEZ MANTEIGA, W. (1996). NN goodness-of-fit tests for linear models. *J. Statist. Plann. Inference* **53**, 75–92.

VAN DER VAART, A.D. (2000). *Asymptotic statistics.* Cambridge University Press, Cambridge series on statistical and probabilistic mathematics.

VAN DER VAART, A.D., AND WELLNER, J.A. (1996). *Weak convergence and empirical processes.* Springer Series in Statistics. Springer-Verlag, New-York.

Yao, F., and Müller, H.G. (2010). Functional quadratic regression. *Biometrika* **97**, 49–64.

Yao, F., and Müller, H.G., and Wang, J-L. (2005). Functional linear regression analysis for longitudinal data.. *Annals of Statistics* **33**, 2873–2903.

Yurinsky, V. (1995). *Sums and Gaussian Vectors.* Springer series in statistics. Probability and its applications. Springer-Verlag, Berlin Heidelberg.

Zheng, J.X. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics* **75**, 263–289.

INSA de RENNES
Service des Formations

AVIS DU JURY SUR LA REPRODUCTION DE LA THESE SOUTENUE

Titre de la thèse:
Contributions à l'analyse statistique des données fonctionnelles

Nom Prénom de l'auteur : SAUMARD MATTHIEU

Membres du jury :
- Monsieur PATILEA Valentin
- Monsieur LEDOUX James
- Monsieur SARDA Pascal
- Monsieur CARDOT Hervé
- Monsieur MAS André

Président du jury : Pascal SARDA

Date de la soutenance : 23 Mai 2013

Reproduction de la these soutenue

☒ Thèse pouvant être reproduite en l'état
☐ Thèse pouvant être reproduite après corrections suggérées

Fait à Rennes, le 23 Mai 2013

Signature du président de jury

Le Directeur,

M'hamed DRISSI